Complementarity Constraints as Nonlinear Equations: Theory and Numerical Experience*

Sven Leyffer
Mathematics and Computer Science Division
Argonne National Laboratory, Argonne, IL 60439, USA.
leyffer@mcs.anl.gov

September 3, 2003 revised February 2, 2005

Abstract

Recently, it has been shown that mathematical programs with complementarity constraints (MPCCs) can be solved efficiently and reliably as nonlinear programs. This paper examines various nonlinear formulations of the complementarity constraints. Several nonlinear complementarity functions are considered for use in MPCC. Unlike standard smoothing techniques, however, the reformulations do not require the control of a smoothing parameter. Thus they have the advantage that the smoothing is *exact* in the sense that Karush-Kuhn-Tucker points of the reformulation correspond to strongly stationary points of the MPCC. A new *exact smoothing* of the well-known min function is also introduced and shown to possess desirable theoretical properties. It is shown how the new formulations can be integrated into a sequential quadratic programming solver, and their practical performance is compared on a range of test problems.

Keywords: Nonlinear programming, SQP, MPCC, complementarity constraints, NCP functions.

AMS-MSC2000: 90C30, 90C33, 90C55, 49M37, 65K10.

1 Introduction

Equilibrium constraints in the form of complementarity conditions often appear as constraints in optimization problems, giving rise to mathematical programs with complementarity constraints (MPCCs). Problems of this type arise in many engineering and economic applications; see the survey [11] and the monographs [24, 26]. The growing collections of test problems [22, 7] indicate that this an important area. MPCCs can be

*Preprint ANL/MCS-P1054-0603

expressed in general as

$$
\begin{aligned}
\text{minimize} \quad & f(x) && (1.1a)\\
\text{subject to} \quad & c_{\mathcal{E}}(x) = 0 && (1.1b)\\
& c_{\mathcal{I}}(x) \geq 0 && (1.1c)\\
& 0 \leq x_1 \ \perp \ x_2 \geq 0, && (1.1d)
\end{aligned}
$$

where $x = (x_0, x_1, x_2)$ is a decomposition of the problem variables into controls $x_0 \in \mathbb{R}^n$ and states $x_1, x_2 \in \mathbb{R}^p$. The equality constraints $c_i(x) = 0,\ i \in \mathcal{E}$ are abbreviated as $c_{\mathcal{E}}(x) = 0$, and similarly $c_{\mathcal{I}}(x) \geq 0$ represents the inequality constraints. The notation $\perp$ represents complementarity and means that either a component $x_{1i} = 0$ or the corresponding component $x_{2i} = 0$.

Clearly, more general complementarity constraints can be included in (1.1) by adding slack variables. Adding slacks does not destroy any properties of the MPCC such as constraint qualification or second-order condition. One convenient way of solving (1.1) is to replace the complementarity conditions (1.1d) by

$$x_1,\ x_2 \geq 0, \quad \text{and} \quad X_1 x_2 \leq 0, \tag{1.2}$$

where X_1 is a diagonal matrix with x_1 along its diagonal. This transforms the MPCC into an equivalent nonlinear program (NLP) and is appealing because it appears to allow standard large-scale NLP solvers to be used to solve (1.1).

Unfortunately, it has been shown [5, 29] that (1.2) violates the Mangasarian-Fromovitz constraint qualification (MFCQ) at *any* feasible point. This failure of MFCQ has a number of unpleasant consequences: The multiplier set is unbounded, the central path fails to exist, the active constraint normals are linearly dependent, and linearizations of the NLP can be inconsistent *arbitrarily close* to a solution. In addition, early numerical experience with (1.2) has been disappointing [2]. As a consequence, solving MPCCs as NLPs has been commonly regarded as numerically unsafe.

Recently, exciting new developments have demonstrated that the gloomy prognosis about the use of (1.2) may have been premature. Standard sequential quadratic programming (SQP) solvers have been used to solve a large class of MPCCs, written as NLPs, reliably and efficiently [16]. This numerical success has motivated a closer investigation of the (local) convergence properties of SQP methods for MPCCs. In [17], it is shown that an SQP method converges locally to strongly stationary points. Anitescu [1] establishes that an SQP method with *elastic mode* converges locally for MPCCs with (1.2). The key idea is to penalize $X_1 x_2 \leq 0$ and consider the resulting NLP, which satisfies MFCQ. Near a strongly stationary point, a sufficiently large penalty parameter can be found, and standard SQP methods converge.

The convergence properties of interior point methods (IPMs) have also received renewed attention. Numerical experiments [3, 28] have shown that IPMs with minor modifications can be applied successfully to solve MPCCs. This practical success has encouraged theoretical studies of the convergence properties of IPMs for MPCCs. Raghunathan and Biegler [27] relax $x_1^T x_2 \leq 0$ by a quantity proportional to the barrier parameter, which is driven to zero. Liu and Sun [23] propose a primal-dual IPM that also relaxes the complementarity constraint.

In this paper, we extend our results of [17] by considering NLP formulations of (1.1) in which the complementarity constraint (1.1d) is replaced by an nonlinear complementarity problem (NCP) function. This gives rise to the following NLP:

$$\begin{aligned}
&\text{minimize} && f(x) && \text{(1.3a)}\\
&\text{subject to} && c_{\mathcal{E}}(x) = 0 && \text{(1.3b)}\\
& && c_{\mathcal{I}}(x) \geq 0 && \text{(1.3c)}\\
& && x_1,\ x_2 \geq 0,\ \Phi(x_{1i}, x_{2i}) \leq 0, && \text{(1.3d)}
\end{aligned}$$

where $\Phi(x_1, x_2)$ is the vector of NCP functions, $\Phi(x_1, x_2) = (\phi(x_{11}, x_{21}), \ldots, \phi(x_{1p}, x_{2p}))^T$, and ϕ is any NCP function introduced in the next section. Problem (1.3) is in general nonsmooth because the NCP functions used in (1.3d) are nonsmooth at the origin. We will show that this nonsmoothness does not affect the local convergence properties of the SQP method.

The use of NCP functions for the solution of MPCCs has been considered in [8, 10], where a sequence of smoothed NCP reformulation is solved. Our contribution is to show that this smoothing is not required. Thus we avoid the need to control the smoothing parameter that may be problematic in practice. Moreover, the direct use of NCP functions makes our approach exact, in the sense that first-order points of the resulting NLP coincide with strongly stationary points of the MPCC. As a consequence we can prove superlinear convergence under reasonable assumptions.

The paper is organized as follows. The next section reviews the NCP functions that will be used in (1.3d) and their pertinent properties. We also introduce new NCP functions shown to possess certain desirable properties. Section 3 shows the equivalence of first-order points of (1.1) and (1.3). Section 4 formally introduces the SQP algorithm for solving MPCCs. The equivalence of the first-order conditions forms the basis of the convergence proof of the SQP method, presented in Section 5. In Section 6, we examine the practical performance of the different NCP functions on the MacMPEC test set [22]. In Section 7 we summarize our work and briefly discuss open questions.

2 NCP Functions for MPCCs

An NCP function is a function $\phi : \mathbb{R}^2 \to \mathbb{R}$ such that $\phi(a, b) = 0$ if and only if $a, b \geq 0$, and $ab \leq 0$. Several NCP functions can be used in the reformulation (1.3). Here, we review some existing NCP functions and introduce new ones that have certain desirable properties.

1. The bilinear form
$$\phi_{BL}(a, b) = ab, \tag{2.4}$$
which is analytic and has the appealing property that its gradient vanishes at the origin (this makes it consistent with strong stationarity, as will be shown later). It is not an NCP function, however, since $\phi_{BL}(a, b) = 0$ does not imply nonnegativity of a, b.

2. The Fischer-Burmeister function [12] is given by
$$\phi_{FB}(a, b) = a + b - \sqrt{a^2 + b^2}. \tag{2.5}$$

It is nondifferentiable at the origin, and its Hessian is unbounded at the origin.

3. The min-function [6] is the nonsmooth function

$$\phi_{\min}(a,b) = \min(a,b). \tag{2.6}$$

It can be written equivalently in terms of the natural residual function [6]:

$$\phi_{NR}(a,b) = \frac{1}{2}\left(a+b-\sqrt{(a-b)^2}\right). \tag{2.7}$$

This function is again nondifferentiable at the origin and along the line $a=b$.

4. The Chen-Chen-Kanzow function [4] is a convex combination of the Fischer-Burmeister function and the bilinear function. For a fixed parameter $\lambda \in (0,1)$, it is defined as

$$\phi_{CCK}(a,b) = \lambda\phi_{FB}(a,b) + (1-\lambda)a_+b_+,$$

where $a_+ = \max(0,a)$. Note that for $a \geq 0$, $a_+ = a$; hence, for any method that remains feasible with respect to the simple bounds,

$$\phi_{CCK}(a,b) = \lambda\phi_{FB}(a,b) + (1-\lambda)\phi_{BL}(a,b) \tag{2.8}$$

holds.

We note that all functions (except for (2.4)) are nondifferentiable at the origin. In addition, the Hessian of the Fischer-Burmeister function is unbounded at the origin. This has to be taken into account in the design of robust SQP methods for MPCCs.

The min-function has the appealing property that linearizations of the resulting NLP (1.3) are consistent sufficiently close to a strongly stationary point (see Proposition 3.6). This property motivates the derivation of smooth approximations of the min-function. The first approximation is obtained by smoothing the equivalent natural residual function (2.7) by adding a term to the square root (which causes the discontinuity along $a=b$). For a *fixed parameter* $\sigma_{NR} > 1/2$, let

$$\phi_{NRs}(a,b) = \frac{1}{2}\left(a+b-\sqrt{(a-b)^2+\frac{ab}{\sigma_{NR}}}\right). \tag{2.9}$$

This smoothing is similar to [6, 10], where a positive parameter $4\mu^2 > 0$ is added to the discriminant. This has the effect that complementarity is satisfied only up to μ^2 at the solution. In contrast, adding the term ab/σ_{NR}, implies that the NCP function remains *exact* in the sense that $\phi_{NRs}(a,b) = 0$ if and only if $a,b \geq 0$ and $ab = 0$ for *any* $\sigma_{NR} > 1/2$. Figure 2 shows the contours of $\phi_{NRs}(a,b)$ for $\sigma_{NR} = 32$ and for the min-function ($\sigma_{NR} = \infty$). An interesting observation is that as $\sigma_{NR} \to \frac{1}{2}$, the smoothed min-function $\phi_{NRs}(a,b)$ becomes the Fischer-Burmeister function (up to a scaling factor).

An alternative way to smooth the natural residual function is to work directly on smoothing the contours of the min-function, which are piecewise constant. The contours can be smoothed by dividing the positive orthant into (for example) three regions as shown in Figure 1. The dashed lines separate the three regions (i) to (iii), and their

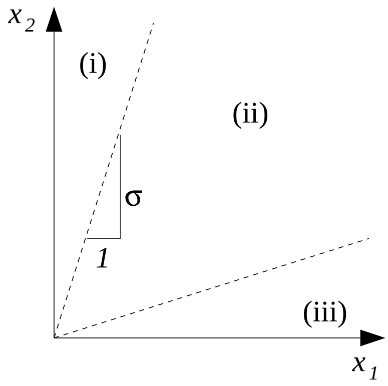

Figure 1: Piecewise regions for smoothing the min-function

slope is $\sigma > 1$ and σ^{-1}, respectively. In regions (i) and (iii), the contours are identical to the min-function. This feature ensures consistency of the linearization. In region (ii), different degrees of smoothing can be applied.

The first smoothed min-function is based on a piecewise linear approximation, given by

$$\phi_{lin}(a,b) \;=\; \begin{cases} b & b \le a/\sigma_l \\ (a+b)/(1+\sigma_l) & a/\sigma_l < b < \sigma_l a \\ a & b \ge \sigma_l a, \end{cases} \tag{2.10}$$

where $\sigma = \sigma_l > 1$ is the parameter that defines the three regions in Figure 1. The idea is that close to the axis, the min-function is used, while for values of a, b that are in the center, the decision as to which should be zero is delayed.

The second smoothed min-function is based on the idea of joining the linear parts in sectors (i) and (iii) with circle segments. This gives rise to the following function,

$$\phi_{qua}(a,b) \;=\; \begin{cases} b & b \le a/\sigma_q \\ \sqrt{\dfrac{(a-\theta)^2+(b-\theta)^2}{(\sigma_q-1)^2}} & a/\sigma_q < b < \sigma_q a \\ a & b \ge \sigma_q a, \end{cases} \tag{2.11}$$

where θ is the center of the circle, depending on a, b, and σ_q and is given by

$$\theta \;=\; \frac{a+b}{2-\frac{(\sigma_q-1)^2}{\sigma_q^2}} + \sqrt{\left(\frac{a+b}{2-\frac{(\sigma_q-1)^2}{\sigma_q^2}}\right)^2 - \frac{a^2+b^2}{2-\frac{(\sigma_q-1)^2}{\sigma_q^2}}}.$$

The contours of both smoothing functions are given in Figure 2. Note that the contours are parallel to the axis in regions (i) and (iii). This fact will be exploited to show that linearizations of the min-function and its two variants remain consistent arbitrarily close to a strongly stationary point. This observation, in effect, establishes a constraint qualification for the equivalent NLP (1.3).

The smoothing also avoids another undesirable property of the min-function: It projects iterates that are far from complementary onto the nearest axis. Close to the axis $a = b$, this projection results in an arbitrary step. Consider, for example, a point $a = 99, b = 101$.

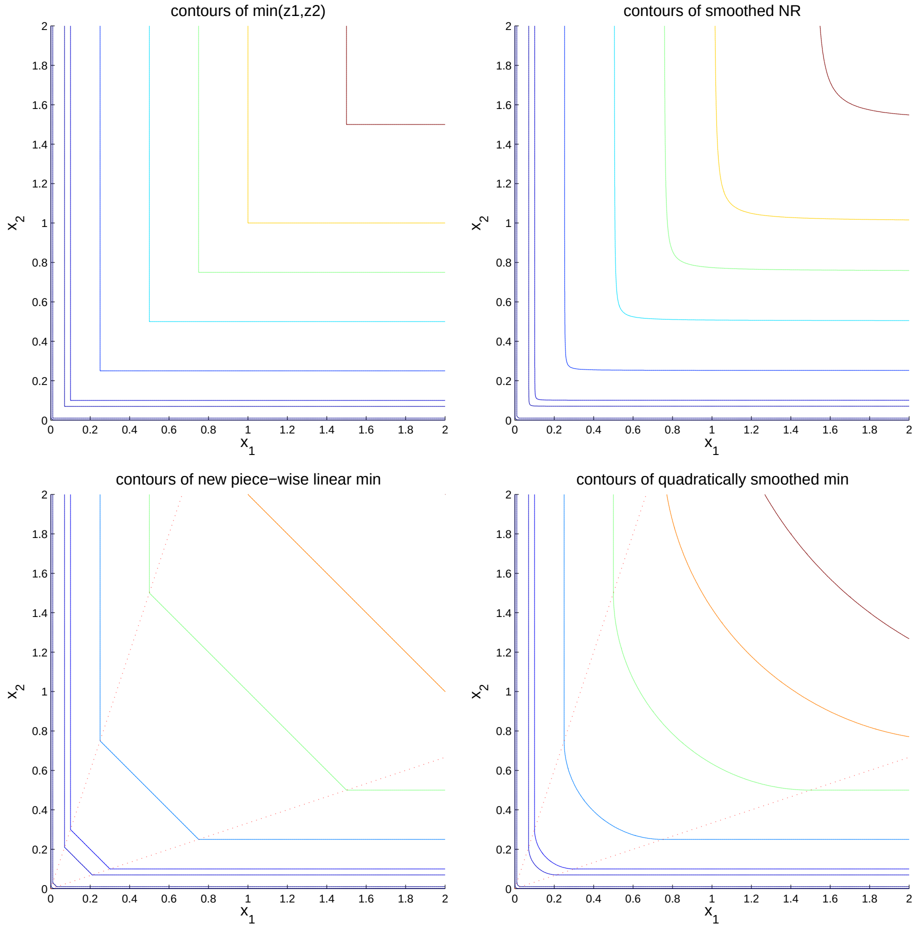

Figure 2: Contours of the min-function, the smoothed natural residual function, the piecewise linear min-function, and the piecewise quadratic with $\sigma_l = \sigma_q = 3$

Linearizing the min-function about this point results in a first-order approximation in which $a = 0, b \geq 0$. In contrast, other NCP functions "delay" this decision and can be viewed as smoothing methods.

3 Equivalence of First-Order Conditions

This section shows that there exists a one-to-one correspondence between strongly stationary points of the MPCC (1.1) and the first-order stationary points of the equivalent NLP (1.3). We start by reviewing MPCC stationarity concepts. Next, we derive some properties of the linearizations of (1.3d) that play a crucial role in the equivalence of first-order conditions.

3.1 Strong Stationarity for MPCCs

The pertinent condition for stationarity for analyzing NLP solvers applied to (1.3) is strong stationarity. The reason is that there exists a relationship between strong stationarity [29] and the Karush-Kuhn-Tucker (KKT) points of (1.3). This relationship has been exploited in [17] to establish convergence of SQP methods for MPCCs formulated as NLPs. Strong stationarity is defined as follows.

Definition 3.1 *A point x is called* strongly stationary *if and only if there exist multipliers λ, $\hat{\nu}_1$, and $\hat{\nu}_2$ such that*

$$\begin{aligned} \nabla f(x) - \nabla c^T(x)\lambda - \begin{pmatrix} 0 \\ \hat{\nu}_1 \\ \hat{\nu}_2 \end{pmatrix} &= 0 \\ c_{\mathcal{E}}(x) &= 0 \\ c_{\mathcal{I}}(x) &\geq 0 \\ x_1,\ x_2 &\geq 0 \\ x_{1j} = 0 \ \ \textit{or} \ \ x_{2j} &= 0 \\ \lambda_{\mathcal{I}} &\geq 0 \\ c_i\lambda_i = x_{1j}\hat{\nu}_{1j} = x_{2j}\hat{\nu}_{2j} &= 0 \\ \textit{if } x_{1j} = x_{2j} = 0 \ \ \textit{then} \ \ \hat{\nu}_{1j} \geq 0 \quad \textit{and} \quad \hat{\nu}_{2j} &\geq 0. \end{aligned} \tag{3.1}$$

Strong stationarity can be interpreted as the KKT conditions of the relaxed NLP (3.2) at a feasible point x. Given two index sets $\mathcal{X}_1,\ \mathcal{X}_2 \subset \{1,\ldots,p\}$ with

$$\mathcal{X}_1 \cup \mathcal{X}_2 \;=\; \{1,\ldots,p\}\,,$$

denote their respective complements in $\{1,\ldots,p\}$ by $\mathcal{X}_1^{\perp}$ and $\mathcal{X}_2^{\perp}$. For any such pair of index sets, define the *relaxed NLP corresponding to the MPCC (1.1)* as

$$\begin{array}{lllll} \underset{x}{\text{minimize}} & f(x) & & & \\ \text{subject to} & c_{\mathcal{E}}(x) = 0 & & & \\ & c_{\mathcal{I}}(x) \geq 0 & & & \\ & x_{1j} = 0 & \forall j \in \mathcal{X}_2^{\perp} & \text{and} & x_{1j} \geq 0 \quad \forall j \in \mathcal{X}_2 \\ & x_{2j} = 0 & \forall j \in \mathcal{X}_1^{\perp} & \text{and} & x_{2j} \geq 0 \quad \forall j \in \mathcal{X}_1. \end{array} \tag{3.2}$$

Concepts such as MPCC constraint qualifications (CQs) and second-order conditions are defined in terms of this relaxed NLP (see, e.g., [17]). Formally, the linear independence constraint qualification (LICQ) is extended to MPCCs as follows:

Definition 3.2 *The MPCC (1.1) is said to satisfy an* MPCC-LICQ *at x if the corresponding relaxed NLP (3.2) satisfies an LICQ.*

Next, a second-order sufficient condition (SOSC) for MPCCs is given. Like strong stationarity, it is related to the relaxed NLP (3.2). Let $\mathcal{A}^*$ denote the set of active constraints of (3.2) and $\mathcal{A}^*_+ \subset \mathcal{A}^*$ the set of active constraints with nonzero multipliers (some could be negative). Let A denote the matrix of active constraint normals, that is,

$$A = \left[\begin{array}{ccccccc} & & & & 0 & & 0 \\ A^*_{\mathcal{E}} & : & A^*_{\mathcal{I}\cap\mathcal{A}^*} & : & I^*_1 & : & 0 \\ & & & & 0 & & I^*_2 \end{array} \right] =: \; [a^*_i]_{i\in\mathcal{A}^*}\,,$$

where $A^*_{\mathcal{I}\cap\mathcal{A}^*}$ are the active inequality constraint normals and

$$I_1^* := [e_i]_{i\in\mathcal{X}_1^*} \quad \text{and} \quad I_2^* := [e_i]_{i\in\mathcal{X}_2^*}$$

are parts of the $p\times p$ identity matrices corresponding to active bounds. Define the set of feasible directions of zero slope of the relaxed NLP (3.2) as

$$S^* = \left\{ s \mid s \neq 0 \ , \ {g^*}^T s = 0 \ , \ {a_i^*}^T s = 0 \ , \ i \in \mathcal{A}^*_+ \ , \ {a_i^*}^T s \geq 0 \ , \ i \in \mathcal{A}^* \backslash \mathcal{A}^*_+ \right\}.$$

The MPCC-SOSC is defined as follows.

Definition 3.3 *A strongly stationary point z^* with multipliers $(\lambda^*, \hat{\nu}_1^*, \hat{\nu}_2^*)$ satisfies the* MPCC-SOSC *if for every direction $s \in S^*$ it follows that $s^T \nabla^2 \mathcal{L}^* s > 0$, where $\nabla^2\mathcal{L}^*$ is the Hessian of the Lagrangian of (3.2) evaluated at $(z^*, \lambda^*, \hat{\nu}_1^*, \hat{\nu}_2^*)$.*

3.2 Linearizations of the NCP Functions

All NCP functions with the exception of the bilinear form are nonsmooth at the origin. In addition, the min-function is also nonsmooth along $a = b$, and the linearized min-function is nonsmooth along $a = \sigma^{-1}b$ and $a = \sigma b$. Luckily, SQP methods converge for a simple choice of subgradient.

We start by summarizing some well-known properties of the gradients of the Fischer-Burmeister function (2.5) for $(a, b) \neq (0, 0)$:

$$\nabla \phi_{FB}(a,b) = \begin{pmatrix} 1 - \dfrac{a}{\sqrt{a^2+b^2}} \\ 1 - \dfrac{b}{\sqrt{a^2+b^2}} \end{pmatrix}.$$

It can be shown that $0 < 1 - \frac{a}{\sqrt{a^2+b^2}} < 2$ for all $(a, b) \neq (0, 0)$. In addition, if $a > 0$ and $b > 0$, it can be shown that

$$\nabla \phi_{FB}(a,0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ and } \nabla \phi_{FB}(0,b) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Similarly, the gradient of the smoothed natural residual function is

$$\nabla \phi_{NRs}(a,b) = \frac{1}{2} \begin{pmatrix} 1 - \dfrac{a - b + \frac{b}{2\sigma}}{\sqrt{(a-b)^2 + \frac{ab}{\sigma}}} \\ 1 - \dfrac{b - a + \frac{a}{2\sigma}}{\sqrt{(a-b)^2 + \frac{ab}{\sigma}}} \end{pmatrix}.$$

For $a > 0$ and $b > 0$, it follows that

$$\nabla \phi_{NRs}(a,0) = \begin{pmatrix} 0 \\ 1 - \frac{1}{4\sigma} \end{pmatrix} \text{ and } \nabla \phi_{NRs}(0,b) = \begin{pmatrix} 1 - \frac{1}{4\sigma} \\ 0 \end{pmatrix}.$$

Despite the fact that the NCP functions are not differentiable everywhere, it turns out that a particular choice of subgradient gives fast convergence for SQP methods. To show equivalence of the first-order conditions in [17], we exploit the fact that $\nabla \phi_{BL}(0,0) = 0$. Fortunately, 0 is a generalized gradient of the other NCP functions, that is, $0 \in \partial\phi(0,0)$. Similarly, we will choose a suitable subgradient for the min-function along $a = b$. With a slight abuse of notation, we summarize the subgradient convention:

Convention 3.4 *The following convention is used for subgradients of the nonsmooth NCP functions:*

1. $\nabla\phi(0,0) = 0$ *for any NCP function.*
2. $\nabla\phi_{\min}(a,a) = (\frac{1}{2}, \frac{1}{2})^T$ *for the min-function for* $a > 0$.
3. $\nabla\phi_{lin}(a, \sigma a) = (0,1)$ *and* $\nabla\phi_{lin}(a, \sigma^{-1} a) = (1,0)$ *for the linearized min-function, for* $a > 0$.

This convention is consistent with the subgradients of the NCP functions and is readily implemented. The most important convention is to ensure that $\nabla\phi(0,0) = 0$ because, otherwise, we would not be able to establish equivalence of first-order conditions. The other conventions could be relaxed to allow other subgradients. The convention on the subgradients also has an important practical implication. We have observed convergence to M-stationary, or even C-stationary points that are not strongly stationary for other choices of $0 \neq v \in \partial\phi(0,0)$. Setting $v = 0 \in \partial\phi(0,0)$ prevents convergence to such spurious stationary points.

It turns out that a straightforward application of SQP to (1.3) is not very efficient in practice. The reason is that the linearization of the complementarity constraint (1.2) together with the lower bounds has no strict interior. Therefore, we relax the linearization of (1.2). Let $0 < \delta, \kappa < 1$ be constants, and consider

$$a \geq 0,\ b \geq 0,\ \phi(\hat{a}, \hat{b}) + \nabla\phi(\hat{a}, \hat{b})^T \begin{pmatrix} a - \hat{a} \\ b - \hat{b} \end{pmatrix} \leq \delta \left(\min(1, \phi(\hat{a}, \hat{b})) \right)^{1+\kappa}. \tag{3.3}$$

Clearly, this is a relaxation of the linearization of (1.2). The following proposition summarizes some useful properties of the linearizations of the NCP functions.

Proposition 3.5 *Let* $\phi(a,b)$ *be one of the functions (2.4)–(2.11). Then it follows that*

1. $a, b \geq 0$ *and* $\phi(a,b) \leq 0$ *is equivalent to* $0 \leq a \perp b \geq 0$.
2. *If* $\hat{a}, \hat{b} \geq 0$ *and* $\hat{a} + \hat{b} > 0$, *then it follows that the perturbed system of inequalities (3.3) is consistent for any* $0 \leq \delta, \kappa \leq 1$. *In addition, if* $\delta > 0$ *and* $\hat{a}, \hat{b} > 0$, *then (3.3) has a nonempty interior for the Fischer-Burmeister function, the bilinear function, and the smoothed natural residual function.*

Proof. Part 1 is obvious. For Part 2, consider each NCP function in turn. For the bilinear function (2.4), it readily follows that $(a,b) = (0,0)$ is feasible in (3.3) because for $\hat{a}, \hat{b} \geq 0$, we get

$$\hat{a}\hat{b} + \nabla\phi_{BL}^T \begin{pmatrix} -\hat{a} \\ -\hat{b} \end{pmatrix} = -\hat{a}\hat{b} \leq 0,$$

and clearly, if $\delta > 0$, there exists a nonempty interior.

Next consider the Fischer-Burmeister function (2.5), for which (3.3) with $\delta = 0$ becomes

$$\left(1 - \frac{\hat{a}}{\sqrt{\hat{a}^2 + \hat{b}^2}}\right) a + \left(1 - \frac{\hat{b}}{\sqrt{\hat{a}^2 + \hat{b}^2}}\right) b \leq 0.$$

Since the terms in the parentheses are positive, it follows that $(a,b) = 0$ is the only point satisfying $a, b \geq 0$ and (3.3). On the other hand, if $\delta > 0$, then the right-hand side of (3.3) is positive, and there exists a nonempty interior of $a, b \geq 0$ and (3.3).

For (2.6) and (2.7), it follows for $\hat{a} < \hat{b}$ that (3.3) becomes $a = 0, b \geq 0$. The conclusion for $\hat{a} > \hat{b}$ follows similarly. If $\hat{a} = \hat{b}$, then (3.3) becomes $\frac{1}{2}a + \frac{1}{2}b \leq \delta \left(\min(1, \phi(\hat{a}, \hat{b}))\right)^{1+\kappa}$, and the results follow.

The result for (2.8) follows from the fact that (2.8) is a linear combination of the Fischer-Burmeister function and (2.4).

To show the result for the smoothed min functions, we observe that for $b \leq a/\sigma$ and $b \geq \sigma a$ the functions are identical to the min-function and the result follows. For $a/\sigma < b < \sigma a$, we consider (2.10) and (2.11) in turn. The linearization of (2.10) is equivalent to $a + b \leq 0$, which implies feasibility. It can also be shown that linearization of (2.11) about any point is feasible at the origin $(a,b) = (0,0)$.

The smoothed natural residual function also has feasible linearizations. For (2.9), (3.3) is equivalent to (using $\sigma = \sigma_{NR}$ to simplify the notation)

$$\phi_{NRs}(\hat{a}, \hat{b}) + \left(1 - \frac{\hat{a} - \hat{b} + \frac{\hat{b}}{2\sigma}}{\sqrt{(\hat{a} - \hat{b})^2 + \frac{\hat{a}\hat{b}}{\sigma}}}\right)(a - \hat{a}) + \left(1 - \frac{\hat{b} - \hat{a} + \frac{\hat{a}}{2\sigma}}{\sqrt{(\hat{a} - \hat{b})^2 + \frac{\hat{a}\hat{b}}{\sigma}}}\right)(b - \hat{b}) \leq 0.$$

Rearranging, we have

$$-\phi_{NRs}(\hat{a}, \hat{b}) + \left(1 - \frac{\hat{a} - \hat{b} + \frac{\hat{b}}{2\sigma}}{\sqrt{(\hat{a} - \hat{b})^2 + \frac{\hat{a}\hat{b}}{\sigma}}}\right) a + \left(1 - \frac{\hat{b} - \hat{a} + \frac{\hat{a}}{2\sigma}}{\sqrt{(\hat{a} - \hat{b})^2 + \frac{\hat{a}\hat{b}}{\sigma}}}\right) b \leq 0.$$

The first term is clearly nonpositive, and it can be shown that the terms multiplying a and b are nonnegative, thus implying consistency and a nonempty interior, even when $\delta = 0$. $\square$

A disadvantage of the functions (2.4), (2.8), and (2.9) is that arbitrarily close to a strongly stationary point, the linearizations may be inconsistent [17]. The next proposition shows that the min-function and its smoothed versions (2.10) and (2.11) do not have this disadvantage.

Proposition 3.6 *Consider (1.3) using any of the min-functions, (2.6), (2.10), or (2.11), and assume that the MPCC-MFCQ holds at a strongly stationary point. Then it follows that the linearization of (1.3) is consistent for all $x_1, x_2 \geq 0$ sufficiently close to this strongly stationary point.*

Proof. Under MPCC-MFCQ, it follows that the linearization of the relaxed NLP (3.2) is consistent in a neighborhood of a strongly stationary point. Now consider the linearization of the min-function near a strongly stationary point, x^* say. For components i, such that $x^*_{1i} = 0 < x^*_{2i}$, it follows for any point x^k sufficiently close to x^* that $0 \leq x^k_{1i} < x^k_{2i}$. Thus, the linearization of the corresponding min-function gives $d_{1i} \leq -x^k_{1i}$. Together with the lower bound $d_{1i} \geq -x^k_{1i}$, this is equivalent to $d_{1i} = -x^k_{1i}$, the linearization of the same component in the relaxed NLP. A similar conclusion holds for components with $x^*_{1i} > 0 = x^*_{2i}$.

Finally, for components i, such that $x^*_{1i} = 0 = x^*_{2i}$, it follows that the origin $x^{k+1}_{1i} = x^{k+1}_{2i} = 0$ is feasible (Proposition 3.5). This point is also feasible for the relaxed NLP.

A similar argument can be made for the smoothed min-functions (2.10) and (2.11) by observing that for $x^*_{1i} = 0 < x^*_{2i}$, there exists a neighborhood where these functions agree with the min-function and for $x^*_{1i} = 0 = x^*_{2i}$, feasibility follows from Proposition 3.5. □

An important consequence of this proposition is that the quadratic convergence proof for MPCCs in [17] can now be applied *without* the assumption that all QP subproblems are consistent. In this sense, Proposition 3.6 implies that the equivalent NLP (1.3) using the min-functions satisfies a constraint qualification.

3.3 NCP Functions and Strong Stationarity

A consequence of the gradient convention is that the gradients of all NCP functions have the same structure. In particular, it follows that for $a, b > 0$

$$\nabla\phi(a,0) = \begin{pmatrix} 0 \\ \tau_a \end{pmatrix} \quad \nabla\phi(0,b) = \begin{pmatrix} \tau_b \\ 0 \end{pmatrix} \quad \nabla\phi(a,b) = \begin{pmatrix} \tau_b \\ \tau_a \end{pmatrix} \quad \text{and} \quad \nabla\phi(0,0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

for some parameters $\tau_a, \tau_b > 0$ that depend on a, b and the NCP function. As a consequence, we can generalize the proof of equivalence of first-order conditions from [17] to all NCP functions from Section 2. Let $\Phi(x_1, x_2)$ denote the vector of functions $\phi(x_{1i}, x_{2i})$. The KKT conditions of (1.3) are that there exist multipliers $\mu := (\lambda, \nu_1, \nu_2, \xi)$ such that

$$\begin{aligned} \nabla f(x) - \nabla c(x)^T\lambda - \begin{pmatrix} 0 \\ \nu_1 \\ \nu_2 \end{pmatrix} + \begin{pmatrix} 0 \\ \nabla_{x_1}\Phi(x_1,x_2)\xi \\ \nabla_{x_2}\Phi(x_1,x_2)\xi \end{pmatrix} &= 0 \\ c_{\mathcal{E}}(x) &= 0 \\ c_{\mathcal{I}}(x) &\geq 0 \\ x_1, x_2 &\geq 0 \\ \Phi(x_1,x_2) &\leq 0 \\ \lambda_{\mathcal{I}} &\geq 0 \\ \nu_1, \nu_2 &\geq 0 \\ \xi &\geq 0 \\ c_i(x)\lambda_i = x_{1j}\nu_{1j} = x_{2j}\nu_{2j} &= 0 \quad . \end{aligned} \tag{3.4}$$

There is also a complementarity condition $\xi^T\Phi(x_1,x_2)=0$, which is implied by feasibility of x_1,x_2 and has been omitted. Note that the choice $\nabla\phi(0,0)=0$ makes (3.4) consistent with strong stationarity, as will be shown next.

Theorem 3.7 *$(x^*,\lambda^*,\hat\nu_1,\hat\nu_2)$ is a strongly stationary point satisfying (3.1) if and only if there exist multipliers $(x^*,\lambda^*,\nu_1^*,\nu_2^*,\xi^*)$ satisfying the KKT conditions (3.4) of the equivalent NLP (1.3). If ϕ is any of the NCP function of Section 2, then*

$$\hat\nu_1 = \nu_1^* - \boldsymbol{\tau}_1\xi^* \tag{3.5a}$$
$$\hat\nu_2 = \nu_2^* - \boldsymbol{\tau}_2\xi^*, \tag{3.5b}$$

where $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ are diagonal matrices of with $\tau_j, j=1,2$ along their diagonals. Moreover, $\tau_{ji}=0$, if $x_{1i}=x_{2i}=0$ and otherwise satisfies the relationship

$$\tau_{1i} = \begin{cases} 1 \;\; \textit{if} \;\; x_{2i}>0 & \textit{for } (2.5),(2.6),(2.7),(2.10),(2.11) \\ 1-\frac{1}{4\sigma} \;\; \textit{if} \;\; x_{2i}>0 & \textit{for } (2.9) \\ x_{2i} & \textit{for } (2.4) \\ \lambda+(1-\lambda)x_{2i} \;\; \textit{if} \;\; x_{2i}>0 & \textit{for } (2.8) \end{cases} \tag{3.6}$$

and

$$\tau_{2i} = \begin{cases} 1 \;\; \textit{if} \;\; x_{1i}>0 & \textit{for } (2.5),(2.6),(2.7),(2.10),(2.11) \\ 1-\frac{1}{4\sigma} \;\; \textit{if} \;\; x_{1i}>0 & \textit{for } (2.9) \\ x_{1i} & \textit{for } (2.4) \\ \lambda+(1-\lambda)x_{1i} \;\; \textit{if} \;\; x_{1i}>0 & \textit{for } (2.8). \end{cases} \tag{3.7}$$

Proof. Note that gradients $\nabla\Phi$ have the same structure for all NCP functions used. Then (3.5) follows by comparing (3.4) and (3.1) and taking the gradients of the NCP functions into account. □

The failure of MFCQ for (1.3) implies that the multiplier set is unbounded. However, this unboundedness occurs in a special way. The multipliers of (1.3) form a ray, similar to [17], and there exists a multiplier of minimum norm, given by

$$\nu_{1i}^* = \max(\hat\nu_{1i},0), \tag{3.8a}$$
$$\nu_{2i}^* = \max(\hat\nu_{2i},0), \tag{3.8b}$$
$$\xi_i^* = -\min\left(\frac{\hat\nu_{1i}}{\tau_{1i}},\frac{\hat\nu_{2i}}{\tau_{2i}},0\right). \tag{3.8c}$$

This implies the following complementarity conditions for the multipliers

$$0 \le \nu_{1i}^* \;\perp\; \xi_i^* \ge 0 \;\text{ and }\; 0 \le \nu_{2i}^* \;\perp\; \xi_i^* \ge 0. \tag{3.9}$$

This multiplier will be referred to as the *minimal, or basic, multiplier*. This term is justified by the observation (to be proved below) that the constraint normals corresponding to nonzero components of the basic multiplier are linearly independent, provided the MPCC satisfies an LICQ.

4 An SQP Algorithm for NCP Functions

This section describes an SQP algorithm for solving (1.3). The algorithm is an iterative procedure that solves a quadratic programming (QP) approximation of (1.3) around the iterate x^k for a step d at each iteration:

$$(QP^k)\begin{cases} \underset{d}{\text{minimize}} & g^{k^T} d + \frac{1}{2} d^T W^k d \\ \text{subject to} & c_{\mathcal{E}}^k + A_{\mathcal{E}}^{k^T} d = 0 \\ & c_{\mathcal{I}}^k + A_{\mathcal{I}}^{k^T} d \geq 0 \\ & x_1^k + d_1 \geq 0 \\ & x_2^k + d_2 \geq 0 \\ & \Phi^k + \nabla_{x_1}\Phi^{k^T} d_1 + \nabla_{x_2}\Phi^{k^T} d_2 \leq \delta \left(\min(1, \Phi^k)\right)^{1+\kappa}, \end{cases}$$

where $\mu^k = (\lambda^k, \nu_1^k, \nu_2^k, \xi^k)$ and $W^k = \nabla^2\mathcal{L}(x^k, \mu^k)$ is the Hessian of the Lagrangian of (1.1):

$$W^k = \nabla^2\mathcal{L}(x^k, \mu^k) = \nabla^2 f(x^k) - \sum_{i \in \mathcal{I} \cup \mathcal{E}} \lambda_i \nabla^2 c_i(x^k).$$

Note that the Hessian W^k does not include entries corresponding to $\nabla^2\Phi$. This omission is deliberate as it avoids numerical difficulties near the origin, where $\nabla^2\phi_{FB}$ becomes unbounded. It will be shown that this does not affect the convergence properties of SQP methods.

The last constraint of (QP^k) is the relaxation of the linearization of the complementarity condition (3.3). We will show that the perturbation does not impede fast local convergence. Formally, the SQP algorithm is defined in Algorithm 1.

Let $k = 0$, x^0 given
while *not optimal* **do**
  Solve (QP^k) for a step d
  Set $x^{k+1} = x^k + d$, and $k = k + 1$

Algorithm 1: Local SQP Algorithm for MPCCs

In practice, we also include a globalization scheme to stabilize SQP. In our case, we use a filter [15] and a trust region to ensure convergence to stationary points [18]. The convergence theory of filter methods allows for three possible outcomes [18, Theorem 1]:

(A) The algorithm terminates at a point that is locally infeasible.

(B) The algorithm converges to a Kuhn-Tucker point.

(C) The algorithm converges to a feasible point at which MFCQ fails.

Clearly, **(B)** cannot happen because (1.3) violates MFCQ at any feasible point. Outcome **(A)** is typically associated with convergence to a local minimum of the norm of the constraint violation and cannot be avoided unless global optimization techniques are used. Therefore, we deal mainly with outcome **(C)** if we apply a filter algorithm to MPCC formulated as NLPs (1.3). The next section presents a local convergence analysis of the SQP algorithm applied to (1.3).

5 Local Convergence of SQP for MPCCs

This section establishes superlinear convergence of SQP methods a strongly stationary point under mild conditions. The notation $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2$ introduced in Theorem 3.7 allows the convergence analysis of all NCP functions to be unified. We note that the presence of the perturbation term $\delta \left(\min(1, \Phi^k)\right)^{1+\kappa}$, with $\kappa < 1$, implies that we cannot obtain quadratic convergence in general.

The convergence analysis is concerned with strongly stationary points. Let x^* be a strongly stationary point, and denote by $\mathcal{A}(x^*)$ the set of active general constraints:

$$\mathcal{A}(x^*) := \{i | c_i(x^*) = 0\}.$$

We also denote the set of active bounds by

$$\mathcal{X}_j(x^*) := \{i | x_{ji} = 0\} \quad \text{for } j = 1, 2$$

and let $\mathcal{D}(x^*) := \mathcal{X}_1(x^*) \cap \mathcal{X}_2(x^*)$ be the set of degenerate indices associated with the complementarity constraint.

Assumptions 5.1 *We make the following assumptions:*

[**A0**] *The subgradients of the NCP functions are computed according to Convention 3.4.*

[**A1**] *The functions f and c are twice Lipschitz continuously differentiable.*

[**A2**] *(1.1) satisfies an MPCC-LICQ.*

[**A3**] *x^* is a strongly stationary point that satisfies an MPCC-SOSC.*

[**A4**] $\lambda_i \neq 0$, $\forall i \in \mathcal{E}^*$, $\lambda_i^* > 0$, $\forall i \in \mathcal{A}^* \cap \mathcal{I}$, *and either* $\nu_{1j}^* > 0$ *and* $\nu_{2j}^* > 0$, $\forall j \in \mathcal{D}^*$.

[**A5**] *The QP solver always chooses a linearly independent basis.*

We note that [**A0**] is readily implemented and that assumption [**A5**] holds for the QP solvers used within `snopt` [20] and `filter` [15]. The most restrictive assumptions are [**A2**] and [**A3**] because they exclude B-stationary points that are not strongly stationary. This fact is not surprising because it is well known that SQP methods typically converge linearly to such B-stationary points.

It is useful to divide the convergence proof into two parts. First, we consider the case where complementarity holds for some iterate k, i.e. $\Phi(x_1^k, x_2^k) = 0$. In this case, the SQP method applied to (1.3) is shown to be equivalent to SQP applied to the relaxed NLP (3.2). In the second part, we assume that $\Phi(x_1^k, x_2^k) > 0$ for all k. Under the additional assumption that all QP approximations remain consistent, superlinear convergence can again be established.

5.1 Local Convergence for Exact Complementarity

In this section we make the additional assumption that

[A6] $\Phi(x_1^k, x_2^k) = 0$ and (x^k, μ^k) is sufficiently close to a strongly stationary point.

Assumption **[A6]** implies that for given index sets $\mathcal{X}_j := \mathcal{X}_j(x^k) := \left\{i | x_{ji}^k = 0\right\}, j = 1, 2$, the following holds:

$$\begin{array}{lll} x_{1j}^k = 0 & & \forall j \in \mathcal{X}_2^\perp \\ x_{2j}^k = 0 & & \forall j \in \mathcal{X}_1^\perp \\ x_{1j}^k = 0 & \text{or} \;\; x_{2j}^k = 0 & \forall j \in \mathcal{D} = \mathcal{X}_1 \cap \mathcal{X}_2. \end{array}$$

In particular, it is not necessary to assume that both $x_{1i}^k = 0$ and $x_{2i}^k = 0$ for $i \in \mathcal{D}^*$. Thus it may be possible that $\mathcal{X}_1 \neq \mathcal{X}_1^*$ (and similarly for $\mathcal{X}_2$). An important consequence of **[A6]** is that $\mathcal{X}_1$, $\mathcal{X}_2$ satisfy

$$\begin{array}{ccccc} \mathcal{X}_1^{*^\perp} & \subset & \mathcal{X}_1^\perp & \subset & \mathcal{X}_1^{*^\perp} \cup \mathcal{D}^* \\ \mathcal{X}_2^{*^\perp} & \subset & \mathcal{X}_2^\perp & \subset & \mathcal{X}_2^{*^\perp} \cup \mathcal{D}^* \\ & & \mathcal{D} & \subset & \mathcal{D}^*, \end{array} \tag{5.1}$$

that is, the indices $\mathcal{X}_1^{*^\perp}$ and $\mathcal{X}_2^{*^\perp}$ of the nondegenerate complementarity constraints have been identified correctly.

Next, it is shown that SQP applied to (1.3) is equivalent to SQP applied to the relaxed NLP (3.2). For a given partition $(\mathcal{X}_1^\perp, \mathcal{X}_2^\perp, \mathcal{D})$, an SQP step for the relaxed NLP (3.2) is obtained by solving the QP

$$(QP_R(x^k)) \left\{ \begin{array}{lllll} \underset{d}{\text{minimize}} & g^{k^T} d + \frac{1}{2} d^T W^k d & & & \\ \text{subject to} & c_{\mathcal{E}}^k + A_{\mathcal{E}}^{k^T} d = 0 & & & \\ & c_{\mathcal{I}}^k + A_{\mathcal{I}}^{k^T} d \geq 0 & & & \\ & d_{1j} = 0 & \forall j \in \mathcal{X}_2^\perp & \text{and} & x_{1j}^k + d_{1j} \geq 0 \quad \forall j \in \mathcal{X}_2 \\ & d_{2j} = 0 & \forall j \in \mathcal{X}_1^\perp & \text{and} & x_{2j}^k + d_{2j} \geq 0 \quad \forall j \in \mathcal{X}_1. \end{array} \right.$$

The following proposition shows that SQP applied to the relaxed NLP converges quadratically and identifies the correct index sets $\mathcal{X}_1^*$ and $\mathcal{X}_2^*$ in one step. Its proof can be found in [17, Proposition 5.2].

Proposition 5.2 *Let* **[A1]**–**[A6]** *hold, and let x^k be sufficiently close to x^*. Consider the relaxed NLP for any index sets $\mathcal{X}_1$, $\mathcal{X}_2$ (satisfying (5.1) by virtue of* **[A6]***). Then it follows that*

1. *there exists a neighborhood U of $(z^*, \lambda^*, \nu_1^*, \nu_2^*)$ and a sequence of iterates generated by SQP applied to the relaxed NLP (3.2), $\{(x^l, \lambda^l, \nu_1^l, \nu_2^l)\}_{l>k}$, that lies in U and converges Q-quadratically to $(x^*, \lambda^*, \nu_1^*, \nu_2^*)$;*
2. *the sequence $\{x^l\}_{l>k}$ converges Q-superlinearly to x^*; and*
3. *$\mathcal{X}_1^l = \mathcal{X}_1^*$ and $\mathcal{X}_2^l = \mathcal{X}_2^*$ for $l > k$.*

Next, it is shown that the QP approximation to the relaxed NLP $(QP_R(x^k))$ and the QP approximation to the NCP formulation (QP^k) generate the same sequence of steps. The next lemma shows that the solution of $(QP_R(x^k))$ is feasible in (QP^k).

Lemma 5.3 *Let Assumptions* [**A1**]–[**A6**] *hold. Then it follows that a step d is feasible in $(QP_R(x^k))$ if and only if it is feasible in (QP^k).*

Proof. $(QP_R(x^k))$ and (QP^k) differ only in the way the complementarity constraint is treated. Hence we need only to prove the equivalence of those constraints. Let $j \in \mathcal{X}_2^\perp$. Then it follows that $x_{1j} = 0$, and $\frac{\partial \Phi^k}{\partial x_{1j}} = \tau_{1j} > 0$, and $\frac{\partial \Phi^k}{\partial x_{1j}} = 0$. Hence, (QP^k) contains the constraints

$$\tau_{1j} d_{1j} \leq 0 \ \text{ and } \ d_{1j} \geq 0 \ \Leftrightarrow \ d_{1j} = 0.$$

Similarly, we can show that the constraints are equivalent for $j \in \mathcal{X}_1^\perp$. Let $j \in \mathcal{D}$. Then it follows that (QP^k) contains the constraints $d_{2j} \geq 0$ and $d_{1j} \geq 0$, which are equivalent to the constraints of $(QP_R(x^k))$. The equivalence of the feasible sets follows because $(\mathcal{X}_1^\perp, \mathcal{X}_2^\perp, \mathcal{D})$ is a partition of $\{1, \ldots, p\}$. □

The next lemma shows that the solution of the two QPs are identical and that the multipliers are related.

Lemma 5.4 *Let Assumptions* [**A1**]–[**A6**] *hold. Let $(\lambda, \hat{\nu}_1, \hat{\nu}_2)$ be the Lagrange multipliers of $(QP_R(x^k, \mathcal{X}))$ (corresponding to a step d). Then it follows that the multipliers of (QP^k), corresponding to the same step d are $\mu = (\lambda, \nu_1, \nu_2, \xi)$, where*

$$\nu_{1i} = \quad \hat{\nu}_{1i} > 0, \ \forall i \in \mathcal{D} \tag{5.2a}$$

$$\nu_{2i} = \quad \hat{\nu}_{2i} > 0, \ \forall i \in \mathcal{D} \tag{5.2b}$$

$$\xi_i = \quad -\min(\frac{\hat{\nu}_{1i}}{\tau_{1i}}, \frac{\hat{\nu}_{2i}}{\tau_{2i}}, 0) \tag{5.2c}$$

$$\nu_{1i} = \hat{\nu}_{1i} - \xi_i \tau_{1i}, \ \forall i \in \mathcal{X}_2^\perp \tag{5.2d}$$

$$\nu_{2i} = \hat{\nu}_{2i} - \xi_i \tau_{2i}, \ \forall i \in \mathcal{X}_1^\perp, \tag{5.2e}$$

where τ_{ji} is given in (3.6–3.7). Conversely, given a solution d and multipliers μ of (QP^k), (5.2) shows how to construct multipliers so that $(d, \lambda, \widehat{\nu_1}, \widehat{\nu_2})$ solves $(QP_R(x^k, \mathcal{X}))$.

Proof. We equate the first-order conditions of $(QP_R(x^k))$ and (QP^k) and obtain

$$g^k + W^k d - A^k \lambda \ = \ \begin{pmatrix} 0 \\ \hat{\nu}_1 \\ \hat{\nu}_2 \end{pmatrix} \ = \ \begin{pmatrix} 0 \\ \nu_1 - \nabla_{x_1} \Phi \xi \\ \nu_2 - \nabla_{x_2} \Phi \xi \end{pmatrix}.$$

We distinguish three cases:

Case 1 ($j \in \mathcal{D}$): It follows from (5.1) that $j \in \mathcal{D}^*$, which implies that $\hat{\nu}_{1j}, \hat{\nu}_{2j} > 0$ for x^k sufficiently close to x^* by assumption [**A4**]. Moreover, $\frac{\partial \Phi}{\partial x_{1j}} = \frac{\partial \Phi}{\partial x_{2j}} = 0$, and hence, $\nu_{1j} = \hat{\nu}_{1j} > 0$, $\nu_{2j} = \hat{\nu}_{2j} > 0$, and $\xi_j = 0$ are valid multipliers for (QP^k).

Case 2 ($j \in \mathcal{X}_1^\perp$): We distinguish two further cases. If $j \in \mathcal{D}^*$, then a similar argument to Case 1 shows that $\nu_{1j} = \hat{\nu}_{1j} > 0$, $\nu_{2j} = \hat{\nu}_{2j} > 0$, and $\xi_j = 0$. On the other hand, if $j \in \mathcal{X}_1^{*^\perp}$, then it follows that $\frac{\partial \Phi}{\partial x_{1j}} = 0$, and $\frac{\partial \Phi}{\partial x_{2j}} = \tau_{2j} > 0$ is bounded away from zero. Thus, $\nu_{1j} = \hat{\nu}_{1j} = 0$, and $\nu_{2j} = \hat{\nu}_{2j} - \tau_{2j}\xi_j$, and we can always choose $\nu_{2j}, \xi_j \geq 0$. We will show later that the QP solver in fact chooses either $\nu_{2j} > 0$, or $\xi_j > 0$.
Case 3 ($j \in \mathcal{X}_2^\perp$) is similar to Case 2. □

Next, it is shown that both QPs have the same solution in a neighborhood of $d = 0$; its proof can be found in [17, Lemma 5.6].

Lemma 5.5 *The solution d of ($QP_R(x^k)$) is the only strict local minimizer in a neighborhood of $d = 0$ and its corresponding multipliers $(\lambda, \hat{\nu}_1, \hat{\nu}_2)$ are unique. Moreover, d is also the only strict local minimizer in a neighborhood of $d = 0$ of (QP^k).*

The next theorem summarizes the results of this section.

Theorem 5.6 *If Assumptions* [**A1**]–[**A6**] *hold, then SQP applied to (1.3) generates a sequence $\{(x^l, \lambda^l, \nu_1^l, \nu_2^l, \xi^l)\}_{l>k}$ that converges Q-quadratically to $\{(x^*, \lambda^*, \nu_1^*, \nu_2^*, \xi^*)\}$ of (3.4), satisfying strong stationarity. Moreover, the sequence $\{x^l\}_{l>k}$ converges Q-superlinearly to x^* and $\Phi(x_1^l, x_2^l) = 0$ for all $l \geq k$.*

Proof. Under Assumptions [**A1**]–[**A4**], SQP converges quadratically when applied to the relaxed NLP (3.2). Lemmas 5.3–5.5 show that the sequence of iterates generated by this SQP method is equivalent to the sequence of steps generated by SQP applied to (1.3). This implies Q-superlinear convergence of $\{x^l\}_{l>k}$. Convergence of the multipliers follows by considering (5.2). Clearly, the multipliers in (5.2a) and (5.2b) converge, as they are just the multipliers of the relaxed NLP, which converge by virtue of Proposition 5.2. Now observe that (5.2c) becomes

$$\xi_i^{k+1} = -\min\left(\frac{\hat{\nu}_{1i}^{k+1}}{\tau_{1i}^{k+1}}, \frac{\hat{\nu}_{2i}^{k+1}}{\tau_{2i}^{k+1}}, 0\right).$$

The right-hand side of this expression converges because $\hat{\nu}_{1i}^{k+1}, \hat{\nu}_{2i}^{k+1}$ converge and the denominators τ_i^{k+1} are bounded away from zero for $i \in \mathcal{X}_1^{*^\perp}, \mathcal{X}_2^{*^\perp}$. Finally, (5.2d) and (5.2e) converge by a similar argument.

$\Phi(x_1^l, x_2^l) = 0$, $\forall l \geq k$, follows from the convergence of SQP for the relaxed NLP (3.2) and the fact that SQP retains feasibility with respect to linear constraints. Assumption [**A4**] ensures that $d_{1j}^k = d_{2j}^k = 0, \forall j \in \mathcal{D}^*$, since $\nu_{1j}^k, \nu_{2j}^k > 0$ for biactive complementarity constraints. Thus SQP will not move out of the corner and stay on the same face. □

5.2 Local Convergence for Nonzero Complementarity

This section shows that SQP converges superlinearly even if complementarity does not hold at the starting point, that is, if $\Phi(x_1^k, x_2^k) > 0$. It is shown in [17] that the QP approximation to (1.3) with $x_{1i}x_{2i} \le 0$ can be inconsistent arbitrarily close to a strongly stationary point. Similar examples can be constructed for the NCP functions in Section 2. Only the min-function and its piecewise smooth variations guarantee feasibility of the QP approximation near a strongly stationary point (see Proposition 3.6).

Note that by virtue of the preceding section, any component for which $\phi(x_{1i}^k, x_{2i}^k) = 0$ can be removed from the complementarity constraints and instead be treated as part of the general constraints, as $\phi(x_{1i}^l, x_{2i}^l) = 0$ for all $l \ge k$. Hence, it can be assumed without loss of generality that $\Phi(x_1^k, x_2^k) > 0$ for all k.

In the remainder of the proof, it is assumed without loss of generality that $\mathcal{X}_1^{*\perp} = \emptyset$, that is, the solution can be partitioned as

$$x_2^* = \begin{pmatrix} x_{21}^* \\ x_{22}^* \end{pmatrix} = \begin{pmatrix} 0 \\ x_{22}^* \end{pmatrix}, \tag{5.3}$$

where $x_{22}^* > 0$, and $x_1^* = 0$ is partitioned in the same way. This simplifies the notation in the proof.

SQP methods can take arbitrary steps when encountering infeasible QP approximations. In order to avoid the issue of infeasibility, the following assumption is made that often holds in practice.

[**A7**] All QP approximations (QP^k) are consistent for x^k sufficiently close to x^*.

This is clearly an undesirable assumption because it is an assumption on the progress of the method. However, Proposition 3.6 shows that [**A7**] holds for the NCP reformulations involving the min-function. In addition, it is shown in [17] that [**A7**] is satisfied for MPCCs with vertical complementarity constraints that satisfy a mixed-P property. Moreover, the use of the perturbation makes it less likely that the SQP method will encounter infeasible QP subproblems.

The key idea behind our convergence result is to show convergence for any "basic" active set. To this end, we introduce the set of active complementarity constraints

$$\mathcal{C}(x) := \{ i \ : \ \phi(x_{1i}, x_{2i}) = 0 \} .$$

Let $\mathcal{I}(x) := \mathcal{I} \cap \mathcal{A}(x)$, and let the basic constraints be

$$\mathcal{B}(x) := \mathcal{E} \cup \mathcal{I}(x) \cup \mathcal{X}_1(x) \cup \mathcal{X}_2(x) \cup \mathcal{C}(x).$$

The set of strictly active constraints (defined in terms of the basic multiplier, μ, see (3.8)) is given by

$$\mathcal{B}_+(x) \ := \ \{ i \in \mathcal{B}(x) \mid \mu_i \ne 0 \} .$$

Moreover, let B_+^k denote the matrix of strictly active constraint normals at $x = x^k$, namely,

$$B_+^k \ := \ \left[a_i^k \right]_{i \in \mathcal{B}_+(x^k)} ,$$

where a_i^k is the constraint normal of constraint $i \in \mathcal{B}_+(x^k)$.

The failure of any constraint qualification at a solution x^* of the equivalent NLP (1.3) implies that the active constraint normals at x^* are linearly dependent. However, the constraint normals corresponding to strictly active constraints are linearly independent, as shown in the following lemma.

Lemma 5.7 *Let Assumptions* **[A1]**–**[A4]** *hold, and let x^* be a solution of the MPCC (1.1). Let $\mathcal{I}^*$ denote the set of* active *inequalities $c_{\mathcal{I}}(x)$, and consider the matrix of active constraint normals at x^*,*

$$B^* = \begin{bmatrix} & & 0 & 0 & 0 \\ A^*_{\mathcal{E}} & A^*_{\mathcal{I}^*} & I & 0 & \begin{pmatrix} 0 \\ -\nabla_{x_{12}}\Phi_2 \end{pmatrix} \\ & & 0 & \begin{bmatrix} I \\ 0 \end{bmatrix} & \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{bmatrix}, \tag{5.4}$$

where we have assumed without loss of generality that $\mathcal{X}_1^{\perp^*} = \emptyset$. The last column is the gradient of the complementarity constraint. Then it follows that B is linearly dependent and that

$$span\langle \begin{bmatrix} 0 \\ I_2 \end{bmatrix} \rangle = span\langle \begin{bmatrix} 0 \\ -\nabla_{x_{12}}\Phi_2 \end{bmatrix} \rangle. \tag{5.5}$$

Moreover, any submatrix of columns of B has full rank provided that it contains $[A^*_{\mathcal{E}}\ A^*_{\mathcal{I}}]$ and a linearly independent set from the columns in (5.5).

Proof. The structure of the gradient of the NCP functions and (5.3) show that (5.5) holds. Thus B^* is linearly dependent. MPCC-LICQ shows that B^* without the columns corresponding to the NCP functions has full rank. By choosing a linearly independent subset from the columns in (5.5), we get a basis. □

Lemma 5.7 shows that the normals corresponding to the basic multiplier are linearly independent despite the fact that the active normals are linearly dependent. The proof shows that in order to obtain a linearly independent basis, any column of $x_{12} = 0$ can be exchanged with the corresponding normal of the complementarity constraint. This matches the observation that the basic multipliers of the simple bounds and the corresponding complementarity constraint are complementary (see (3.9)).

Next, it is shown that for x^k sufficiently close to x^*, if both the normals corresponding to $x_{1i} \geq 0$ and $\phi(x_{1i}, x_{2i}) \leq 0$ are active, then at the next iteration exact complementarity holds for that component and $\phi(x^l_{1i}, x^l_{2i}) = 0$ and for all subsequent iterations by virtue of Lemma 5.3. Thus, the QP solver cannot continue to choose a basis that is increasingly ill-conditioned.

Lemma 5.8 *Let Assumptions* **[A1]**–**[A5]** *hold, and let x^k be sufficiently close to x^*. Partition the NCP function $\Phi = (\Phi_1, \Phi_2)^T$ in the same way as x_1, x_2 in (5.3). Consider the matrix of active constraint normals at x^k,*

$$B = \left[\begin{array}{cc|cc|cc} & & 0 & 0 & 0 & 0 \\ A^k_{\mathcal{E}} & A^k_{\mathcal{I}} & \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} & & \begin{bmatrix} -\nabla_{x_{11}}\Phi_1 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ -\nabla_{x_{12}}\Phi_2 \end{bmatrix} \\ & & & \begin{bmatrix} I \\ 0 \end{bmatrix} & \begin{bmatrix} -\nabla_{x_{21}}\Phi_1 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ -\nabla_{x_{22}}\Phi_2 \end{bmatrix} \end{array}\right].$$

Then it follows that the columns corresponding to the matrix $\nabla_x\Phi_2$ have the structure $(0, 0, -\tau, 0, -\epsilon)^T$, where $\tau = \mathcal{O}(1)$ and $\epsilon > 0$ is small. If the optimal basis of (QP^k)

contains both a column i of $x_{1i} \geq 0$ and $\phi(x_{1i}, x_{2i}) \leq 0$, then it follows that

$$x_{1i}^k > 0 \;\; and \;\; x_{1i}^{k+1} x_{2i}^{k+1} = 0.$$

Moreover, there exists $c > 0$ such that

$$\| \left(x^{k+1}, \mu^{k+1}\right) - (x^*, \mu^*) \| \; \leq \; c \, \| \left(x^k, \mu^k\right) - (x^*, \mu^*) \|. \tag{5.6}$$

Proof. The first part follows by observing that for x^k close to x^*, $x_{12} \geq 0$ is small and $x_{22} = \mathcal{O}(1)$, which implies the form of the columns. Exchanging them with the corresponding columns of $x_{12} \geq 0$ results in a nonsingular matrix by Lemma 5.7. The second part follows from the nonsingularity assumption [**A5**] (if $x_{1i}^k = 0$, then the basis would be singular) and the fact that if the column corresponding to $x_{1i} \geq 0$ is basic, then $x_{1i}^{k+1} = x_{1i}^k + d_{1i} = 0$ holds.
The third part follows by observing that Assumptions [**A2**] and [**A3**] imply that the relaxed NLP satisfies an LICQ and a SOSC. Hence, the basis B without the final column gives a feasible point close to x^k. Denote this solution by $(\hat{x}, \hat{\mu})$, and let the corresponding step be denoted by $\hat{d}$. Clearly, if this step also satisfies the linearization of the complementarity constraint, that is, if

$$\Phi^k + \nabla_{x_1} \Phi^{k^T} \hat{d}_1 + \nabla_{x_2} \Phi^{k^T} \hat{d}_2 \leq 0,$$

then (5.6) follows by second-order convergence of SQP for the relaxed NLP. If, on the other hand,

$$\Phi^k + \nabla_{x_1} \Phi^{k^T} \hat{d}_1 + \nabla_{x_2} \Phi^{k^T} \hat{d}_2 > 0,$$

then the SQP step of the relaxed NLP is not feasible in (QP^k). In this case consider the following decomposition of the SQP step. Let

$$\hat{d}^n \; = \; \left(\begin{array}{c} 0 \\ \hat{d}_1 \\ \left(\begin{array}{c} \hat{d}_{21} \\ 0 \end{array} \right) \end{array} \right) \; = \; \left(\begin{array}{c} 0 \\ -x_1^k \\ \left(\begin{array}{c} -x_{21}^k \\ 0 \end{array} \right) \end{array} \right)$$

be the normal component, and let $\hat{d}^t := \hat{d} - \hat{d}^n$ be the tangential component. Then it follows that the step of (QP^k) satisfies $d^k = \hat{d}^n + \sigma \hat{d}^t$ for some $\sigma \in [0, 1]$, and the desired bound on the distance follows from the convergence of $\hat{d}$. $\Box$

Thus, if both the normals corresponding to $\phi(x_{1i}, x_{2i}) \leq 0$ and $x_{1i} \geq 0$ are basic, then $x_{1i}^{k+1} x_{2i}^{k+1} = 0$ for a point close to x^*. This component can then be removed from the complementarity constraint, as Lemma 5.3 shows that $x_{1i}^{k+l} x_{2i}^{k+l} = 0$ for all $l \geq 1$. In the remainder we can therefore concentrate on the case that $x_{1i}^k x_{2i}^k > 0$ for all iterates k. Next, it is shown that for x^k sufficiently close to x^*, the basis at x^k contains the equality constraints $\mathcal{E}$ and the active inequality constraints $\mathcal{I}^*$.

Lemma 5.9 *Let x^k be sufficiently close to x^*, and let Assumptions* [**A1**]–[**A5**] *and* [**A7**] *hold. Then it follows that the optimal basis B of (QP^k) contains the normals $A_{\mathcal{E}}^k$ and $A_{\mathcal{I}^*}^k$.*

Proof. This follows by considering the gradient of (QP^k),

$$0 \;=\; \nabla f^k + W^k d^k - \nabla c^{k^T} \lambda^{k+1} - \begin{pmatrix} 0 \\ \nu_1^{k+1} - \xi^{k+1} \nabla_{x_1} \Phi^k \\ \nu_2^{k+1} - \xi^{k+1} \nabla_{x_2} \Phi^k \end{pmatrix},$$

where W^k is the Hessian of the Lagrangian. For x^k sufficiently close to x^*, it follows from **[A4]** that $\lambda_i^{k+1} \neq 0$ for all $i \in \mathcal{E} \cup \mathcal{I}^*$. □

Thus, as long as the QP approximations remain consistent, the optimal basis of (QP^k) will be a subset of B satisfying the conditions in Lemma 5.8. The key idea is to show that for any such basis, there exists an equality constraint problem for which SQP converges quadratically. Since there is only a finite number of basis, this implies convergence for SQP.

We now introduce the *reduced NLP*, which is an equality constraint NLP corresponding to a basis with properties as in Lemma 5.8. Assume that x^* can be partitioned as in (5.3), and define the reduced NLP as

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & c_{\mathcal{E}}(x) = 0 \\ & c_{\mathcal{I}^*}(x) = 0 \\ & x_{11} = 0 \\ & x_{21} = 0 \\ & x_{1i} = 0 \quad \text{or} \quad \Phi(x_{1i}, x_{2i}) = 0 \quad \forall i \in \mathcal{X}_2^\perp, \end{array}$$

where the last constraint means that either $x_{1i} = 0$ or $\Phi(x_{1i}, x_{2i}) = 0$ but not both are present in the reduced NLP. Note that according to (5.3), $\mathcal{X}_1^\perp = \emptyset$. The key idea will be to relate the reduced NLP to a basis satisfying the conditions of Lemma 5.8. Next, it is shown that any reduced NLP satisfies an LICQ and an SOCS.

Lemma 5.10 *Let Assumptions* **[A1]**–**[A4]** *and* **[A7]** *hold. Then it follows that any reduced NLP satisfies LICQ and SOSC.*

Proof. Lemma 5.8 and the fact that either $x_{1i} = 0$ or $\Phi(x_{1i}, x_{2i}) = 0$ are active shows that the normals of the equality constraints of each reduced NLP are linearly independent. The SOSC follows from the MPCC-SOSC and the observation that the MPCC and the reduced NLP have the same null-space. □

Thus, applying SQP to the reduced NLP results in second-order convergence. Next, we observe that any nonsingular basis B corresponds to a reduced NLP. Unfortunately, relaxing the complementarity constraints in (QP^k) means that second-order convergence does not follow directly. However, the particular form of perturbation allows a superlinear convergence result to be established.

Proposition 5.11 *Let Assumptions* **[A1]**–**[A4]** *and* **[A7]** *hold. Then it follows that an SQP method that relaxes the complementarity as in* (QP^k) *converges superlinearly to* x^* *for any reduced NLP.*

Proof. Assume that $\delta = 0$, so that no perturbation is used. Lemma 5.10 shows that the reduced NLP satisfy LICQ and SOSC and, therefore, convergence of SQP follows. In particular, it follows that for a given reduced NLP corresponding to a basis $\mathcal{B}$, there exists a constant $c_{\mathcal{B}} > 0$ such that

$$\| \left(x^{k+1}, \mu^{k+1}\right) - (x^*, \mu^*) \| \ \leq \ c_{\mathcal{B}} \| \left(x^k, \mu^k\right) - (x^*, \mu^*) \|^2. \tag{5.7}$$

If the right-hand side of the complementarity constraint is perturbed (i.e., $\delta > 0$), then consider the Newton step corresponding to the QP approximation of the relaxed NLP about x^k. In particular, this step satisfies $d_N^k = -x_1^k$, and it follows that the perturbation is $o(\|d_N\|)$, where d_N is the Newton step. Hence, superlinear convergence follows using the Dennis-Moré characterization theorem (e.g., [13, Theorem 6.2.3]). ☐

We note that the SQP method based on (QP^k) ignores the curvature corresponding to $\phi(x_{12}, x_{22}) = 0$. However, it is easy to extend the proof of Proposition 5.11 to allow $\nabla^2\Phi$ to be included. The key idea is to show that the limit of the projected Hessian of $\nabla\Phi^*$ is zero. Letting Z^k be a basis of the nullspace of (QP^k), we need to show that $\lim_{k\to\infty} Z^k \nabla^2 \Phi^* = 0$, which implies superlinear convergence (see, e.g., [13, Chapter 12.4]). It can be shown that the Hessian of the NCP functions is unbounded in the nullspace of the active constraints of (QP^k).

Summarizing the results of this section, we obtain the following theorem.

Theorem 5.12 *Let Assumptions* **[A1]**–**[A5]** *and* **[A7]** *hold. Then it follows that SQP applied to the NLP formulation (1.3) of the MPCC (1.1) converges superlinearly near a solution* (x^*, μ^*).

Proof. Proposition 5.11 shows that SQP converges superlinearly for any possible choice of basis $\mathcal{B}$, and Assumption **[A7]** shows that (QP^k) is consistent and remains consistent. Therefore, there exists a basis for which superlinear convergence follows. Thus for each basis,

$$\lim_{k\to\infty} \frac{\|(x^{k+1}, \mu^{k+1}) - (x^*, \mu^*)\|}{\|(x^k, \mu^k) - (x^*, \mu^*)\|} = 0$$

follows. Since there are a finite number of bases, this condition holds independent of the basis and SQP converges superlinearly. ☐

5.3 Discussion of Proofs

Several interesting observations arise from the convergence proofs of the preceding two sections. The curvature of the complementarity constraint $\Phi(x_1, x_2)$ can be ignored without losing fast local convergence. This fact is not surprising because the complementarity constraint

$$0 \leq x_1 \ \perp \ x_2 \geq 0$$

has zero curvature at any feasible point with $x_{1i} + x_{2i} > 0$. At the origin, on the other hand, the curvature is infinite. However, in this case the curvature does not affect convergence, as the reduced Hessian is zero.

If the min-function (2.6) or its piecewise smooth variants (2.10) or (2.11) are used, then the proof simplifies, as near a strongly stationary point, $\nabla\Phi_{x_2} = 0$. In addition, the linearizations are consistent even without the perturbation (3.3) and convergence follows from the convergence of the relaxed NLP. This fact can be interpreted as a constraint qualification for the NCP formulations using (2.6) or (2.10) or (2.11) at strongly stationary points.

The conclusions and proofs presented in this section also carry through for linear complementarity constraints but *not* for general nonlinear complementarity constraints. The reason is that the implication

$$x_{1i}^k x_{2i}^k = 0 \;\Rightarrow\; x_{1i}^{k+1} x_{2i}^{k+1} = 0 \tag{5.8}$$

holds for linear complementarity problems but *not* for nonlinear complementarity problems because in general, an SQP method would move off a nonlinear constraint. This is one reason for the introduction of slacks to deal with more general complementarity constraints. In addition, (5.8) can be made to hold in *inexact arithmetic* by taking care of handling simple bounds appropriately. The same is not true if one expression is a linear equation.

6 Numerical Results

This section describes our experience with an implementation of the different NCP formulation of the MPCC (1.1) in our sequential quadratic programming solver. Our SQP method promotes global convergence through the use of a *filter*. The filter accepts a trial point whenever the objective or the constraint violation is improved compared with all previous iterates [14, 15, 18].

6.1 Preliminaries

The solver includes an AMPL [19] interface that introduces slacks to formulate general complementarity constraints in the form (1.1) and handles the reformulation to the NLP (1.3) automatically. The interface also computes the derivatives of the NCP functions and relaxes the linearizations according to (3.3). A user can choose between the various formulations and set parameters such as δ, κ by passing options to the solver.

The test problems come from MacMPEC [22], a collection of some 150 MPCC test problems [16] from a variety of backgrounds and sizes. The numerical tests are performed on a PC with an Intel Pentium 4 processor with 2.5 GHz and 512 KB RAM running Red Hat Linux version 7.3. The AMPL solver interface is compiled with the Intel C++ compiler version 6.0, and the SQP/MPCC solver is compiled with the Intel Fortran Compiler version 6.0.

Not all 150 problems in MacMPEC are included in this experiment. We have deliberately left out a number of 32×32 discretizations of the incidence set identification and packaging problems. These problems are similar to one another (a small number of them are included) but take a long time to run. This is especially true for the formulations that do not lump the complementarity constraint. In this sense, the results would have been even better for the formulation using the scalar product form.

To determine reasonable values for the various parameters introduced in the definition of the NCP functions, we run a small representative selection of MPCC problems. The overall performance is not very sensitive to a particular parameter choice. No attempt was made to "optimize" the parameter values; rather, we were interested in determining default values that would work well. Table 1 displays the default parameter values.

Table 1: Default parameter values for numerical experiments.

Parameter	Description	Default
δ	relaxation of linearization in (3.3)	0.1
κ	relaxation of linearization in (3.3)	1.0
σ_{NR}	smoothing of natural residual (2.9)	32.0
λ	Chen-Chen-Kanzow parameter (2.8)	0.7
σ_l	slope of linearized min-function (2.10)	4.0
σ_q	slope of quadratic min-function (2.11)	2.0

While the number of parameters may appear unreasonably large, each formulation requires only three parameters to be set. The choice of $\lambda = 0.7$ also agrees with [25], where $\lambda = 0.8$ is suggested. Note that since $\delta = 0.1$, the Chen-Chen-Kanzow function is relaxed further.

Care has to be taken when computing the smoothed natural residual function (2.9); it can be affected by cancellation error, as the following example illustrates. Suppose $a = 10^4$ and $b = 10^{-4}$ and that single-precision arithmetic is used. Then it follows that

$$2\phi_{NRs}(a, b) = (10^4 + 10^{-4}) - \sqrt{(10^4 - 10^{-4})^2 + \frac{1}{\sigma_{NR}}} \overset{\text{float}}{\simeq} 10^4 - \sqrt{10^8} = 0,$$

that is cancellation errors causes (2.9) to declare an infeasible point complementary. This situation can be avoided by employing the same trick used in reformulating the Fischer-Burmeister function [25], giving rise to

$$\phi_{NRs}(a, b) = \frac{1}{2} \frac{\left(\frac{4\sigma_{NR}-1}{\sigma_{NR}}\right)}{a + b + \sqrt{(a-b)^2 + \frac{ab}{\sigma_{NR}}}}. \tag{6.1}$$

Derivative values can be computed in a similarly stable fashion.

6.2 Performance Plots and Results

Results are provided in two forms. The performance plots [9] in Figures 3 and 4 show the relative performance of each formulation in terms of iteration count and CPU time. These plots can be interpreted as follows. For every solver s and every problem p, the ratio of the number of iterations (or CPU time) of solver p on problem s over the fastest solve for problem s is computed and the base 2 logarithm is taken,

$$\log_2 \left(\frac{\#\ \mathrm{iter}(s, p)}{\mathrm{best\_iter}(p)} \right).$$

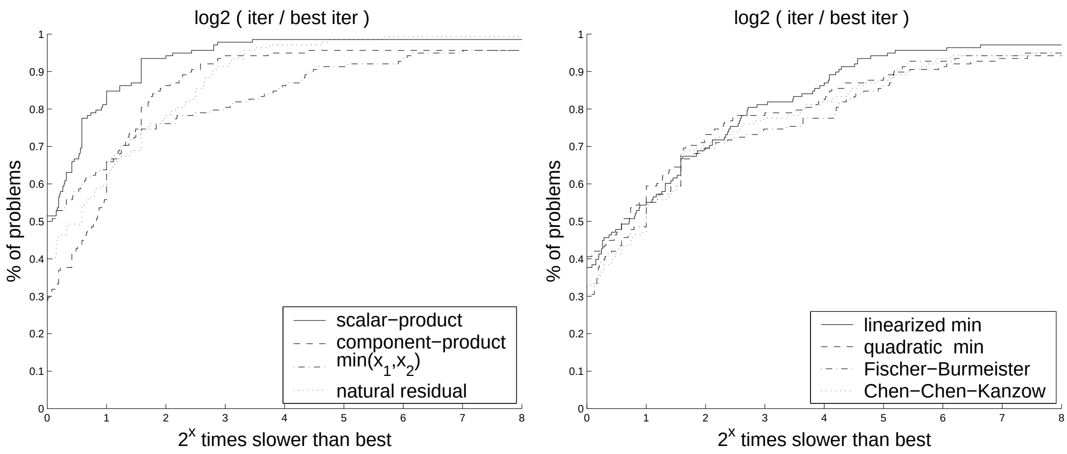

Figure 3: Performance (iterations) plots for different NCP formulations

By sorting these ratios in ascending order for every solver, the resulting plots can be interpreted as the probability that a given solver solves a problem within a certain multiple of the fastest solver.

Failures (see next section) are handled by setting the iteration count and the CPU time to a large number. This strategy ensures that the robustness can also be obtained from the performance plots. The percentage of MPCC problems solved is equivalent to the right asymptote of the performance line for each solver.

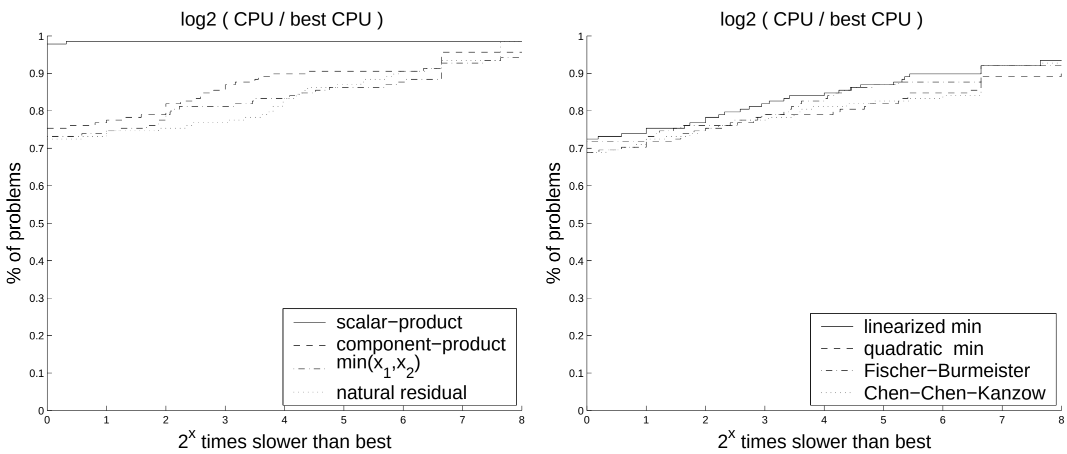

Figure 4: Performance (CPU time) plots for different NCP formulations

6.3 Failures of the NCP Formulations

Solving MPCCs as NLPs is surprisingly robust. We observe very few failures, even though many problems are known to violate the assumptions made in this paper. Even the worst NCP formulation failed only on eight problems. Below, we list the problems that failed together with the reason for the failure.

The NLP solver can fail in three ways. The first failure mode occurs when the trust-region radius becomes smaller than the solver tolerance (`1E-6`) and no further progress can be made. This is referred to in the table below as “TR too small.” Such a failure often happens at a solution where the KKT error cannot be reduced to sufficient accuracy. The second failure mode occurs if the QP solver detects inconsistent linearizations near a feasible point. This is referred to as “infeasible QP.” Note that the fact that MPCCs violate MFCQ implies that linearizations can become inconsistent arbitrarily close to a feasible point. Third, “iter. limit” refers to failures in which the solver reached its iteration limit (1000) without confirming optimality. The following failures were reported:

1. Scalar product form $x_1^T x_2 \leq 0$ — 2 failures

   TR too small : tollmpec1
   infeasible QP : design-cent-3

2. Bilinear form $x_{1i} x_{2i} \leq 0$ — 5 failures

   infeasible QP : design-cent-3, incid-set1c-32, pack-rig2c-32, pack-rig2p-16
   iter. limit : bem-milanc30-s

3. min-function $\min(x_{1i}, x_{2i}) \leq 0$ — 6 failures

   TR too small : ex9.2.2
   infeasible QP : pack-comp1p-8, pack-comp1p-16
   iter. limit : pack-comp2p-8, pack-comp2p-16, qpec-200-2

4. Linearized min-function (2.10) — 4 failures

   TR too small : jr2, qpec-200-3
   infeasible QP : bem-milanc30-s
   iter. limit : qpec-200-2

5. Quadratically smoothed min-function (2.11) — 8 failures

   TR too small : jr2
   infeasible QP : incid-set2c-32
   iter. limit : ex9.2.2, gauvin, incid-set1c-32, qpec-100-4, qpec-200-1, qpec-200-3

6. Fischer-Burmeister function (2.5) — 7 failures

   infeasible QP : design-cent-3, ralphmod
   iter. limit : pack-comp1c-8, pack-rig1-16, pack-rig1c-16, pack-rig2-16,
   : pack-rig2c-16

7. Smoothed natural residual function (2.9) — 1 failures

   TR too small : bem-milanc30-s

8. Chen-Chen-Kanzow function (2.8) — 5 failures

   infeasible QP : pack-comp1p-8, qpec-200-3, pack-comp1c-8, pack-rig2p-16
   iter. limit : bem-milanc30-s

This list contains some problems known not to have strongly stationary limit points. For instance, ex9.2.2, ralph1, and scholtes4 have B-stationary solutions that are not strongly stationary. Problem gauvin has a global minimum at a point where the lower-level problem fails a constraint violation, so the formulation as an MPCC is not appropriate.

In the tests, two problems also gave rise to IEEE errors in the AMPL function evaluations, specifically the Chen-Chen-Kanzow function on `pack-rig1-16` and `pack-rig1c-32`. Since this type of error is caused not by the method but by the model, they are not counted in the errors.

6.4 Interpretation of the Results

The results confirm that solving MPCCs as NLPs is very robust. In particular, the scalar product and the smoothed natural residual function are very robust, solving all but two problems and one problem, respectively.

The results for the min-function, on the other hand, are disappointing. Recall that these functions are theoretically attractive because they do not require an additional assumption to be made on the feasibility of QP approximations. This property makes the number of failures (6/4/8) for the min-function and its smoothed variants disappointing.

The best results in terms of performance and robustness were obtained for the scalar product formulation and the smoothed natural residual function. The performance plots in Figures 3 and 4 clearly show that these formulations are superior. In particular, the scalar product function is significantly faster than any other approach.

The formulation using $x_1^T x_2$ has two main advantages that may explain its superiority. First, it introduces only a single additional constraint, which reduces the size of the NLP to be solved. Moreover, this formulation requires less storage for the QP basis factors. Second, by lumping the complementarity conditions, the formulation allows a certain degree of nonmonotonicity in the complementarity error of each individual $x_{1i}x_{2i}$ and reduces the overall complementarity error, $x_1^T x_2$, only.

The worst results in terms of both robustness and efficiency are obtained for the Fischer-Burmeister function and the quadratically smoothed min-function. These formulations fail on seven and eight problems, respectively and are significantly slower than the other formulations. The Chen-Chen-Kanzow function improves on the Fischer-Burmeister function. This observation is not surprising because ϕ_{CCK} is a convex combination of the Fischer-Burmeister function and the more successful bilinear formulation. The worse behavior of ϕ_{FB} might be due to the fact that its linearized feasible region is smaller than for the bilinear form. This is also supported by the type of failures that can be observed for the Fischer-Burmeister function, which has many infeasible QP terminations.

Analyzing the solution characteristics of the scalar product form, we observe that only four problems have a large value of ξ. This fact shows that the SQP method converges to strongly stationary points for the remaining problems, as a bounded complementarity multiplier is equivalent to strong-stationarity (Theorem 3.7). The four problems for which ξ is unbounded are ex9.2.2, ralph1, ralphmod, and scholtes4. The last problem is known to violate an MPCC-MFCQ at its only stationary point, and the limit is B-stationary but not strongly stationary, and SQP converges linearly for this problem [17].

In addition, it can be observed that the complementarity error is exactly zero at most solutions. The reasons for this behavior are as follows:

1. Complementarity occurs only between variables. Thus, if a lower bound is in the active set, then the corresponding residual can be set to zero even in inexact arithmetic.

2. Many problems in the test set have a solution where $\xi = 0$. This indicates that the complementarity constraint $x_1^T x_2 \leq 0$ is locally redundant. Hence, exact complementarity is achieved as soon as the SQP method identifies the correct active set.

3. Our QP solver resolves degeneracy by making nearly degenerate constraints exactly degenerate and then employing a recursive procedure to remove degeneracy. This process of making nearly degenerate constraints exactly degenerate forces exact complementarity. Consider any nondegenerate index for which $x_{2i}^* > 0 = x_{1i}^*$, and assume that $x_{1i}^k > 0$ is small. The QP solver resolves the "near" degeneracy between the lower bound $x_{1i} \geq 0$ and the complementarity constraint by perturbing x_{1i} to zero. Thus exact complementarity is achieved.

This behavior is reassuring and makes the NLP approach to MPCCs attractive from a numerical standpoint.

7 Conclusions

Mathematical programs with complementarity constraints (MPCCs) are an emerging area of nonlinear optimization. Until recently researchers had assumed that the inherent degeneracy of MPCCs makes the application of standard NLP solvers unsafe. In this paper we show how MPCCs can be formulated as NLPs using a range of so-called NCP functions. Two new smoothed min-functions are introduced that exhibit desirable theoretical properties comparable to a constraint qualification.

In contrast to other smoothing approaches, the present formulations are exact in the sense that KKT points of the reformulated NLP correspond to strongly stationary points of the MPCC. Thus there is no need to control a smoothing parameter, which may be problematic.

It is shown that SQP methods exhibit fast local convergence near strongly stationary points under reasonable assumptions. This behavior is observed in practice on a large range of MPCC problems. The numerical results favor a lumped formulation in which all complementarity constraints are lumped into a single constraint. A new smoothed version of the min-function is also shown to be very robust and efficient. On the other hand, results for other standard NCP functions such as the Fischer-Burmeister function are disappointing.

The use of the simple bounds in the reformulation of complementarity (1.2) allows an alternative NLP formulation of the MPCC (1.1). This formulation lumps the nonlinear NCP functions into a single constraint, similar to $x_1^T x_2 \leq 0$. Thus, an alternative NLP is given by

$$\begin{array}{ll}\text{minimize} & f(x) \\ \text{subject to} & c_{\mathcal{E}}(x) = 0 \\ & c_{\mathcal{I}}(x) \geq 0 \\ & x_1,\ x_2 \geq 0, \\ & e^T \Phi(x_1, x_2) \leq 0.\end{array} \tag{7.1}$$

It is straightforward to see, that (7.1) is equivalent to (1.1). The convergence proof is readily extended to this formulation. We note that (7.1) has several advantages over (1.3). It reduces the number of constraints in the NLP. Moreover, our experience indicates that the lumped version of the bilinear form, $x_1^T x_2 \leq 0$, often performs better than the separate version using $x_{1i} x_{2i} \leq 0$. One reason may be that the lumped version allows nonmonotone changes in the complementarity residual in individual variable pairs as long as the overall complementarity is reduced.

Some open questions remain. One question concerns the global convergence of SQP methods from arbitrary starting points. Any approach to this question must take into account the globalization scheme and, in addition, provide powerful feasibility restoration. A related question is whether SQP methods can avoid convergence to spurious stationary points. Such points are sometimes referred to as C-stationary points even though they allow the existence of trivial first-order descent direction. At present, we believe that current SQP methods cannot avoid convergence to C-stationary points. Any attempt to avoid C-stationarity is likely to require algorithmic modifications.

Acknowledgments

This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Contract W-31-109-ENG-38.

A Detailed Results: Iteration Counts

Name	$x_1^T x_2$	(2.4)	(2.6)	(2.10)	(2.11)	(2.5)	(2.9)	(2.8)
bard1	3	4	9	13	2	25	3	8
bard1m	3	4	9	13	2	4	3	7
bard2	1	1	1	1	1	1	1	1
bard2m	1	1	1	1	1	1	1	1
bard3	4	4	4	4	4	4	4	4
bard3m	4	4	4	4	4	4	4	4
bar-truss-3	10	9	9	9	9	9	9	9
bem-milanc30-s	62	1000	655	111	245	144	410	1000
bilevel1	2	3	2	3	3	4	3	4
bilevel2	7	2	1	2	5	3	1	2
bilevel3	7	6	6	6	6	6	6	6
bilin	2	6	1	3	3	3	5	3
dempe	58	58	58	58	58	94	58	58
design-cent-1	4	4	4	4	4	4	4	4
design-cent-2	31	21	37	37	29	32	32	60
design-cent-3	191	164	173	173	173	217	185	163
design-cent-4	3	4	3	3	3	4	3	4
desilva	2	2	2	2	2	2	2	2
df1	2	2	2	2	2	2	2	2
ex9.1.1	1	2	1	1	1	3	2	3
ex9.1.10	1	1	1	1	1	1	1	1
ex9.1.2	2	3	1	3	3	3	3	3
ex9.1.3	3	3	1	3	4	3	3	3
ex9.1.4	2	2	2	2	2	2	2	2
ex9.1.5	3	3	1	3	3	3	3	3
ex9.1.6	3	5	2	2	2	4	4	6
ex9.1.7	3	3	1	3	3	3	3	3
ex9.1.8	1	1	1	1	1	1	1	1
ex9.1.9	3	3	2	3	8	3	3	3
ex9.2.1	3	4	6	13	6	8	3	8
ex9.2.2	22	22	76	71	1000	238	3	180
ex9.2.3	1	1	1	1	1	1	1	1
ex9.2.4	3	2	2	2	2	2	2	2
ex9.2.5	7	7	1	17	32	35	4	7
ex9.2.6	3	2	1	1	2	2	1	2
ex9.2.7	3	4	6	13	6	8	3	8
ex9.2.8	3	3	1	1	1	4	3	3
ex9.2.9	3	3	1	3	3	3	3	3
flp2	3	3	1	1	1	3	3	1
flp4-1	3	2	2	2	2	2	2	2
flp4-2	3	2	2	2	2	2	2	2
flp4-3	3	2	2	2	2	2	2	2
flp4-4	3	2	2	2	2	2	2	2
gauvin	3	9	71	71	1000	54	7	6

Name	$x_1^T x_2$	(2.4)	(2.6)	(2.10)	(2.11)	(2.5)	(2.9)	(2.8)
gnash10	8	8	7	7	7	8	7	8
gnash11	8	8	7	7	7	8	7	8
gnash12	9	8	8	8	8	8	8	8
gnash13	13	9	10	10	9	10	10	11
gnash14	10	10	9	9	9	13	10	11
gnash15	18	18	41	11	11	9	10	27
gnash16	16	14	26	12	10	45	11	14
gnash17	17	17	10	10	9	11	10	15
gnash18	15	19	55	73	10	184	11	128
gnash19	10	19	10	8	8	18	14	25
hakonsen	10	10	12	12	10	10	10	10
hs044-i	6	4	2	2	4	4	2	4
incid-set1-16	33	139	78	120	493	85	175	66
incid-set1-8	34	35	56	56	51	42	73	65
incid-set1c-16	34	89	89	93	168	69	109	86
incid-set1c-32	37	309	102	155	1000	127	304	161
incid-set1c-8	39	32	43	38	48	35	67	43
incid-set2-16	19	37	35	35	35	33	24	33
incid-set2-8	48	19	18	18	18	18	18	18
incid-set2c-16	37	36	40	35	305	27	71	32
incid-set2c-32	31	87	71	122	489	71	308	88
incid-set2c-8	24	20	27	23	52	29	25	27
jr1	1	1	1	1	1	1	1	1
jr2	7	7	61	66	114	22	3	18
kth1	1	1	1	1	1	1	1	1
kth2	2	2	2	2	2	2	2	2
kth3	4	5	67	67	67	3	2	4
liswet1-050	1	1	1	1	1	1	1	1
liswet1-100	1	1	1	1	1	1	1	1
liswet1-200	1	1	1	1	1	1	1	1
nash1	3	2	1	1	1	2	1	2
outrata31	8	8	7	7	7	8	7	7
outrata32	8	9	8	8	8	9	8	8
outrata33	7	8	7	7	7	8	7	8
outrata34	6	7	6	6	6	7	6	7
pack-comp1-16	20	39	751	64	37	12	68	12
pack-comp1-8	8	30	152	66	24	36	16	36
pack-comp1c-16	5	38	358	76	40	15	50	15
pack-comp1c-32	13	2	787	238	217	50	344	50
pack-comp1c-8	8	19	68	40	14	40	18	41
pack-comp1p-16	45	72	895	344	81	52	197	31
pack-comp1p-8	53	64	274	219	87	36	200	172
pack-comp2-16	43	49	442	42	38	44	81	35
pack-comp2-8	8	26	10	18	11	8	10	8
pack-comp2c-16	15	23	336	76	30	15	17	15
pack-comp2c-32	7	34	901	193	178	45	175	42
pack-comp2c-8	6	11	18	15	14	6	13	6

Name	$x_1^T x_2$	(2.4)	(2.6)	(2.10)	(2.11)	(2.5)	(2.9)	(2.8)
pack-comp2p-16	32	64	1000	190	142	36	232	48
pack-comp2p-8	60	57	1000	104	77	34	171	58
pack-rig1-16	64	56	81	120	178	1000	90	206
pack-rig1-8	7	10	25	13	17	145	13	148
pack-rig1c-16	11	43	15	57	53	458	19	548
pack-rig1c-32	18	238	42	302	369	99	181	107
pack-rig1c-8	6	8	13	13	10	139	9	142
pack-rig1p-16	28	48	56	164	490	97	118	59
pack-rig1p-8	14	16	22	25	60	144	29	147
pack-rig2-16	7	11	21	42	119	1000	30	421
pack-rig2-8	10	16	10	36	38	254	62	253
pack-rig2c-16	6	11	13	67	96	1000	34	421
pack-rig2c-32	11	71	31	187	222	57	551	55
pack-rig2c-8	6	12	6	15	33	254	23	253
pack-rig2p-16	10	38	79	367	436	301	86	309
pack-rig2p-8	20	16	18	46	89	197	20	196
portfl1	5	7	4	21	6	76	6	84
portfl2	4	6	3	43	8	108	5	162
portfl3	4	6	3	8	5	6	10	6
portfl4	4	4	5	7	5	50	8	48
portfl6	4	6	3	4	5	68	8	66
qpec1	3	2	2	2	2	2	2	2
qpec-100-1	7	34	114	112	251	253	43	300
qpec-100-2	7	24	47	137	427	219	44	43
qpec-100-3	6	20	121	137	713	256	27	105
qpec-100-4	5	9	103	497	1000	176	42	78
qpec2	2	2	1	1	1	2	1	2
qpec-200-1	10	24	87	25	1000	363	38	343
qpec-200-2	10	33	1000	1000	888	182	114	79
qpec-200-3	11	20	160	267	1000	377	62	357
qpec-200-4	5	13	78	95	862	92	34	89
ralph1	27	27	70	70	70	368	5	181
ralph2	11	21	1	1	1	168	3	179
ralphmod	7	37	25	46	114	178	48	21
scholtes1	4	3	3	3	3	3	3	3
scholtes2	2	2	2	2	2	2	2	2
scholtes3	4	6	67	67	67	1	1	1
scholtes4	26	28	71	74	74	239	6	181
scholtes5	1	1	1	1	1	1	1	1
sl1	1	1	1	1	1	1	1	1
stackelberg1	4	4	4	4	4	4	4	4
tap-09	21	23	17	18	18	12	11	23
tap-15	28	19	18	12	18	20	19	20
tollmpec	10	36	22	24	20	79	135	128
tollmpec1	10	50	20	28	24	379	139	108
water-FL	272	237	235	279	333	256	263	356
water-net	131	114	109	125	137	126	190	114

B Problem Characteristics

This appendix lists the problem characteristics obtained with the scalar product formulation. The headings in Appendix B are explained in Table 2. The definition of the degree of degeneracy d_1, d_2, d_m is taken from [21]. The corresponding columns refer to first-level degeneracy, d_1, second-level degeneracy, d_2, and mixed-degeneracy, d_m.

Table 2: Headings for tables in Appendix B.

Heading	Description
name	problem name in MacMPEC
n	number of variables (excluding slacks)
m	number of constraints (excluding complementarity)
p	number of complementarity constraints
n_{NLP}	number of variables after slacks added
k	dimension of nullspace at the solution
d_1	number of indices i with $\lambda_i = c_i = 0$
d_2	number of indices i with $x_{1i} = x_{2i} = 0$
d_m	number of indices i with $x_{1i} = x_{2i} = 0$ and ($\nu_{1i} = 0$ or $\nu_{2i} = 0$)
compl	complementarity error ($x_1^T x_2$)
ξ	multiplier of the complementarity constraint $x_1^T x_2 \leq 0$

name	n	m	p	n_{NLP}	k	d_1	d_2	d_m	compl	ξ
bard1	5	4	3	8	0	3	0	0	0.00	0.762
bard1m	6	4	3	9	0	4	0	0	0.00	0.762
bard2	12	9	3	15	0	2	1	0	0.00	0.00
bard2m	12	9	3	15	0	2	1	0	0.00	0.00
bard3	6	5	1	7	0	2	0	0	0.00	0.00
bard3m	6	5	3	9	0	2	0	0	0.00	1.09
bar-truss-3	35	34	6	35	0	13	0	0	0.00	1.45
bem-milanc30-s	3436	3433	1464	3436	1	1745	1	0	0.00	954.
bilevel1	10	9	6	12	0	6	0	0	0.00	0.150
bilevel2	16	13	8	20	1	5	0	0	0.294E-10	0.00
bilevel3	11	10	3	11	0	5	0	0	0.00	1.09
bilin	8	7	6	14	0	4	0	0	0.00	22.0
dempe	3	2	1	4	0	0	0	0	0.00	0.571E-05
design-cent-1	12	11	3	15	0	6	0	0	0.00	2.17
design-cent-2	13	15	3	16	0	11	0	0	0.00	0.00
design-cent-3	15	11	3	18	0	1	0	1	0.00	0.313E-01
design-cent-4	22	20	8	30	1	12	0	0	0.00	0.845
desilva	6	4	2	8	0	2	0	2	0.00	0.00
df1	2	3	1	3	1	1	0	1	0.00	0.00
ex9.1.1	13	12	5	13	0	4	0	0	0.00	0.00
ex9.1.10	11	9	3	11	0	5	0	2	0.00	0.00
ex9.1.2	8	7	2	8	0	4	0	0	0.00	0.00
ex9.1.3	23	21	6	23	0	14	0	1	0.00	3.20
ex9.1.4	8	7	2	8	0	3	0	1	0.00	0.00
ex9.1.5	13	12	5	13	0	8	0	2	0.00	10.0
ex9.1.6	14	13	6	14	0	6	0	1	0.00	1.56
ex9.1.7	17	15	6	17	0	8	0	1	0.00	5.00
ex9.1.8	11	9	3	11	0	5	0	2	0.00	0.00
ex9.1.9	12	11	5	12	0	5	0	1	0.00	0.444
ex9.2.1	10	9	4	10	0	6	0	1	0.00	0.762
ex9.2.2	9	8	3	9	0	3	0	1	0.183E-12	0.386E+07
ex9.2.3	14	13	4	14	0	5	1	0	0.00	0.00
ex9.2.4	8	7	2	8	0	3	0	0	0.00	1.00
ex9.2.5	8	7	3	8	0	3	0	0	0.00	6.00
ex9.2.6	16	12	6	16	2	4	0	2	0.168E-10	0.500
ex9.2.7	10	9	4	10	0	6	0	1	0.00	0.762
ex9.2.8	6	5	2	6	0	3	0	1	0.00	0.500
ex9.2.9	9	8	3	9	0	7	0	0	0.100E-06	0.00
flp2	4	2	2	6	1	2	0	1	0.00	0.987
flp4-1	80	60	30	110	0	30	0	0	0.00	0.00
flp4-2	110	110	60	170	0	60	0	0	0.00	0.00
flp4-3	140	170	70	210	0	70	0	0	0.00	0.00
flp4-4	200	250	100	300	0	100	0	0	0.00	0.00
gauvin	3	2	2	5	0	1	0	0	0.00	0.250

name	n	m	p	n_{NLP}	k	d_1	d_2	d_m	compl	ξ
gnash10	13	12	8	13	1	0	0	0	0.00	0.142
gnash11	13	12	8	13	1	0	0	0	0.00	0.918E-01
gnash12	13	12	8	13	1	0	0	0	0.00	0.397E-01
gnash13	13	12	8	13	1	0	0	0	0.00	0.149E-01
gnash14	13	12	8	13	1	0	0	0	0.00	0.199E-02
gnash15	13	12	8	13	0	3	0	0	0.00	7.65
gnash16	13	12	8	13	0	3	0	0	0.00	1.95
gnash17	13	12	8	13	1	4	0	0	0.00	1.67
gnash18	13	12	8	13	1	4	0	0	0.00	12.7
gnash19	13	12	8	13	0	2	0	0	0.00	2.80
hakonsen	9	8	4	9	0	2	0	0	0.00	0.390
hs044-i	20	14	10	26	0	7	0	1	0.00	5.69
incid-set1-16	485	491	225	485	0	232	0	5	0.00	0.00
incid-set1-8	117	119	49	117	0	54	0	4	0.00	0.00
incid-set1c-16	485	506	225	485	1	233	1	5	0.00	0.00
incid-set1c-32	1989	2034	961	1989	4	165	20	0	0.00	0.00
incid-set1c-8	117	126	49	117	0	59	0	4	0.00	0.00
incid-set2-16	485	491	225	710	3	212	13	0	0.00	0.00
incid-set2-8	117	119	49	166	5	42	7	0	0.00	0.00
incid-set2c-16	485	506	225	710	0	218	12	0	0.00	0.00
incid-set2c-32	1989	2034	961	2950	2	937	24	0	0.00	0.00
incid-set2c-8	117	126	49	166	2	46	6	0	0.00	0.00
jr1	2	1	1	3	1	0	0	0	0.00	0.00
jr2	2	1	1	3	0	0	0	0	0.00	2.00
kth1	2	1	1	2	0	0	1	0	0.00	0.00
kth2	2	1	1	2	1	0	0	0	0.00	0.00
kth3	2	1	1	2	0	0	0	0	0.00	1.00
liswet1-050	152	103	50	202	1	52	0	0	0.00	0.00
liswet1-100	302	203	100	402	1	102	0	0	0.00	0.00
liswet1-200	602	403	200	802	1	202	0	0	0.00	0.00
nash1	6	4	2	8	0	4	0	0	0.00	0.00
outrata31	5	4	4	9	0	0	1	0	0.00	0.164
outrata32	5	4	4	9	1	0	0	0	0.00	0.168
outrata33	5	4	4	9	1	1	0	0	0.00	0.714
outrata34	5	4	4	9	1	1	0	0	0.00	2.07
pack-comp1-16	467	511	225	692	3	268	0	2	0.00	0.00
pack-comp1-8	107	121	49	156	0	113	0	0	0.414E-06	0.00
pack-comp1c-16	467	526	225	692	1	269	0	1	0.00	0.00
pack-comp1c-32	1955	2138	961	2916	3	1108	0	2	0.00	0.00
pack-comp1c-8	107	128	49	156	0	120	0	0	0.414E-06	0.00
pack-comp1p-16	467	466	225	692	5	223	2	0	0.00	0.00
pack-comp1p-8	107	106	49	156	0	83	0	0	0.00	0.00
pack-comp2-16	467	511	225	692	5	268	0	2	0.00	0.00
pack-comp2-8	107	121	49	156	5	62	0	2	0.00	0.00
pack-comp2c-16	467	526	225	692	4	268	0	2	0.00	0.00
pack-comp2c-32	1955	2138	961	2916	16	1058	0	2	0.00	0.00
pack-comp2c-8	107	128	49	156	1	62	0	2	0.00	0.00

name	n	m	p	n_{NLP}	k	d_1	d_2	d_m	compl	ξ
pack-comp2p-16	467	466	225	692	13	223	2	0	0.00	0.00
pack-comp2p-8	107	106	49	156	1	47	2	0	0.00	0.00
pack-rig1-16	380	379	158	485	7	208	0	0	0.00	0.00
pack-rig1-8	87	86	32	108	6	47	0	0	0.00	0.00
pack-rig1c-16	380	394	158	485	4	206	0	0	0.00	0.00
pack-rig1c-32	1622	1652	708	2087	2	928	0	0	0.763E-06	0.00
pack-rig1c-8	87	93	32	108	5	47	0	0	0.00	0.00
pack-rig1p-16	445	444	203	550	3	229	2	0	0.00	0.00
pack-rig1p-8	105	104	47	126	5	50	2	0	0.00	0.00
pack-rig2-16	375	374	149	480	1	203	0	0	0.622E-06	0.00
pack-rig2-8	85	84	30	106	5	43	0	0	0.00	0.00
pack-rig2c-16	375	389	149	480	1	219	0	0	0.484E-06	0.00
pack-rig2c-32	1580	1610	661	2045	0	912	0	0	0.240E-06	0.00
pack-rig2c-8	85	91	30	106	2	45	0	0	0.00	0.00
pack-rig2p-16	436	435	194	541	0	215	1	0	0.00	0.00
pack-rig2p-8	103	102	45	124	6	49	2	0	0.00	0.00
portfl1	87	25	12	87	6	6	0	0	0.00	0.897
portfl2	87	25	12	87	0	7	0	0	0.00	0.682
portfl3	87	25	12	87	13	6	0	0	0.00	31.1
portfl4	87	25	12	87	16	5	0	0	0.00	114.
portfl6	87	25	12	87	13	5	0	0	0.00	55.0
qpec1	30	20	20	40	0	10	10	0	0.00	0.00
qpec-100-1	105	102	100	205	0	74	3	0	0.00	10.1
qpec-100-2	110	102	100	210	0	58	4	0	0.00	191.
qpec-100-3	110	104	100	210	0	35	2	0	0.00	4.45
qpec-100-4	120	104	100	220	0	61	4	0	0.00	15.3
qpec2	30	20	20	40	0	0	10	0	0.00	0.667
qpec-200-1	210	204	200	410	0	153	2	0	0.00	158.
qpec-200-2	220	204	200	420	0	118	2	0	0.00	3.42
qpec-200-3	220	208	200	420	0	48	6	0	0.00	35.5
qpec-200-4	240	208	200	440	0	133	7	0	0.00	7.95
ralph1	2	1	1	3	0	0	0	0	0.471E-12	0.486E+06
ralph2	2	1	1	2	1	0	0	0	0.313E-06	2.00
ralphmod	104	100	100	204	0	79	2	0	0.00	0.357E+08
scholtes1	3	1	1	4	0	1	0	0	0.00	0.00
scholtes2	3	1	1	4	0	0	1	0	0.00	0.00
scholtes3	2	1	1	2	0	0	0	0	0.00	1.00
scholtes4	3	3	1	3	0	0	0	0	0.161E-13	0.525E+07
scholtes5	3	2	2	3	2	0	0	0	0.00	0.00
sl1	8	5	3	11	0	5	0	1	0.00	0.00
stackelberg1	3	2	1	3	1	0	0	0	0.00	0.00
tap-09	86	68	32	118	8	29	0	2	0.00	0.687E-07
tap-15	194	167	83	277	0	169	0	27	0.00	57.1
tollmpec	2403	2376	1748	4151	1	1489	1	88	0.00	2.35
tollmpec1	2403	2376	1748	4151	0	2402	0	86	0.00	0.00
water-FL	213	160	44	213	3	46	0	0	0.00	0.163E+04
water-net	66	50	14	66	2	15	0	0	0.00	0.00

References

[1] M. Anitescu. On solving mathematical programs with complementarity constraints as nonlinear programs. Preprint ANL/MCS-P864-1200, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, 2000.

[2] J. F. Bard. Convex two-level optimization. *Mathematical Programming*, 40(1):15–27, 1988.

[3] H. Benson, D. F. Shanno, and R. V. D. Vanderbei. Interior-point methods for nonconvex nonlinear programming: Complementarity constraints. Technical Report ORFE 02-02, Princeton University, Operations Research and Financial Engineering, 2002.

[4] B. Chen, X. Chen, and C. Kanzow. A penalized Fischer-Burmeister NCP-function. *Mathematical Programming*, 88(1):211–216, June 2000.

[5] Y. Chen and M. Florian. The nonlinear bilevel programming problem: Formulations, regularity and optimality conditions. *Optimization*, 32:193–209, 1995.

[6] B. Chen and P. T. Harker. Smooth approximations to nonlinear complementarity problems. *SIAM J. Optimization*, 7(2):403–420, 1997.

[7] S. P. Dirkse. MPEC world. Webpage, GAMS Development Corp., `www.gamsworld.org/mpec/`, 2001.

[8] S. P. Dirkse, M. C. Ferris, and A. Meeraus. Mathematical programs with equilibrium constraints: Automatic reformulation and solution via constraint optimization. Technical Report NA-02/11, Oxford University Computing Laboratory, July 2002.

[9] E. D. Dolan and J. Moré. Benchmarking optimization software with cops. Technical Report MCS-TM-246, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, November 2000.

[10] F. Facchinei, H. Jiang, and L. Qi. A smoothing method for mathematical programs with equilibrium constraints. *Mathematical Programming*, 85:107–134, 1999.

[11] M. C. Ferris and J. S. Pang. Engineering and economic applications of complementarity problems. *SIAM Review*, 39(4):669–713, 1997.

[12] A. Fischer. A Newton-type method for positive semi-definite linear complementarity problems. *J. Optimization Theory and Applications*, 86:585–608, 1995.

[13] R. Fletcher. *Practical Methods of Optimization, 2^{nd} edition.* John Wiley, Chichester, 1987.

[14] R. Fletcher, N. I. M. Gould, S. Leyffer, Ph. L. Toint, and A. Wächter. Global convergence of trust-region SQP-filter algorithms for general nonlinear programming. *SIAM J. Optimization*, 13(3):635–659, 2002.

[15] R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Mathematical Programming*, 91:239–270, 2002.

[16] R. Fletcher and S. Leyffer. Numerical experience with solving MPECs as NLPs. Numerical Analysis Report NA/210, Department of Mathematics, University of Dundee, Dundee, UK, 2002.

[17] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes. Local convergence of SQP methods for mathematical programs with equilibrium constraints. Numerical Analysis Report NA/209, Department of Mathematics, University of Dundee, Dundee, UK, May 2002.

[18] R. Fletcher, S. Leyffer, and Ph. L. Toint. On the global convergence of a filter-SQP algorithm. *SIAM J. Optimization*, 13(1):44–59, 2002.

[19] R. Fourer, D. M. Gay, and B. W. Kernighan. *AMPL: A Modelling Language for Mathematical Programming*. Books/Cole—Thomson Learning, 2 edition, 2003.

[20] P. E. Gill, W. Murray, and M. A. Saunders. SNOPT: An SQP algorithm for large–scale constrained optimization. *SIAM J. Optimization*, 12(4):979–1006, 2002.

[21] H. Jiang and D. Ralph. QPECgen, a MATLAB generator for mathematical programs with quadratic objectives and affine variational inequality constraints. Technical Report, University of Melbourne, Department of Mathematics, December 1997.

[22] S. Leyffer. MacMPEC: AMPL collection of MPECs. Webpage, `www.mcs.anl.gov/~leyffer/MacMPEC/`, 2000.

[23] X. Liu and J. Sun. Generalized stationary points and an interior point method for mathematical programs with equilibrium constraints. *Mathematical Programming B*, 101(1):231–261, 2004.

[24] Z.-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Cambridge, UK, 1996.

[25] T. S. Munson, F. Facchinei, M. C. Ferris, A. Fischer, and C. Kanzow. The semismooth algorithm for large scale complementarity problems. *INFORMS J. Computing*, 13(4):294–311, 2001.

[26] J. Outrata, M. Kocvara, and J. Zowe. *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*. Kluwer Academic Publishers, Dordrecht, 1998.

[27] A. Raghunathan and L. T. Biegler. Barrier methods for mathematical programs with complementarity constraints (MPCCs). Technical Report, Carnegie Mellon University, Department of Chemical Engineering, Pittsburgh, PA, December 2002.

[28] A. Raghunathan and L. T. Biegler. MPEC formulations and algorithms in process engineering. Technical Report, Carnegie Mellon University, Department of Chemical Engineering, Pittsburgh, PA, 2002.

[29] H. Scheel and S. Scholtes. Mathematical program with complementarity constraints: Stationarity, optimality and sensitivity. *Mathematics of Operations Research*, 25:1–22, 2000.